

Supplemental Materials for “Nonprogressive Diffusion on Social Networks: Approximation and Applications”

Yunduan Lin, Heng Zhang, Renyu Zhang, Zuo-Jun Max Shen

Appendix A: Supporting Arguments for Section 2

A.1. Supporting Arguments for Proposition 1

In this Appendix, we provide the proofs for the theoretical results in Section 2, focusing primarily on the proof of Proposition 1. Before proceeding with that, we present the following lemma and its proof, which will be useful in establishing Proposition 1.

LEMMA EC.1. *Given an arbitrary agent $i \in V$, at least one of the following values*

$$(a) \prod_{\mathbf{y} \in \{0,1\}^{|V|}} \mathbb{P}(Y_i(t) = 1 | \mathbf{Y}(t-1) = \mathbf{y}) \text{ or } (b) \prod_{\mathbf{y} \in \{0,1\}^{|V|}} \mathbb{P}(Y_i(t) = 0 | \mathbf{Y}(t-1) = \mathbf{y}),$$

are positive.

Proof of Lemma EC.1: When random noise $\epsilon_i(t)$ is not bounded on either side, it is obvious that the statement holds. In the following, we only consider the situation when $\epsilon_i(t)$ is with support on some bounded interval $[\underline{\epsilon}, \bar{\epsilon}]$. We will show by contradiction.

If (a) is not positive, then there exists $\mathbf{y} \in \{0,1\}^{|V|}$ such that $\mathbb{P}(Y_i(t) = 1 | \mathbf{Y}(t-1) = \mathbf{y}) = 0$, which means $F_\epsilon\left(-v_i - \beta \frac{\sum_{j \in \mathcal{N}_i} y_j}{d_i}\right) = 1$. Hence, we have

$$-v_i - \inf_{\mathbf{y} \in \{0,1\}^{|V|}} \beta \frac{\sum_{j \in \mathcal{N}_i} y_j}{d_i} = -v_i \geq \underline{\epsilon}.$$

Consequently, we can derive the following inequality

$$-v_i - \sup_{\mathbf{y} \in \{0,1\}^{|V|}} \beta \frac{\sum_{j \in \mathcal{N}_i} y_j}{d_i} = -v_i - \beta > -v_i - \frac{1}{L} \geq -v_i - (\bar{\epsilon} - \underline{\epsilon}) \geq \underline{\epsilon},$$

where the first inequality follows from Assumption 2, the second inequality follows Assumption 1, that is, $|F_\epsilon(\bar{\epsilon}) - F_\epsilon(\underline{\epsilon})| \leq L|\bar{\epsilon} - \underline{\epsilon}|$. As a direct result, for any $\mathbf{y} \in \{0,1\}^{|V|}$, $F_\epsilon(-v_i - \beta \frac{\sum_{j \in \mathcal{N}_i} y_j}{d_i}) \geq F_\epsilon(\underline{\epsilon}) > 0$ and thus the value of (b) is positive.

In conclusion, for each agent $i \in V$, at least one of (a) and (b) have positive value. \square

Proof of Proposition 1: Our goal is to show that the MC has only one recurrent communication class and is aperiodic, regardless of whether it is irreducible.

We first prove that there is only one recurrent communication class. To do this, we construct the vector \mathbf{y}' as follows

$$y'_i = \begin{cases} 0 & \text{when } \prod_{\mathbf{y} \in \{0,1\}^{|V|}} \mathbb{P}(Y_i(t) = 1 | \mathbf{Y}(t-1) = \mathbf{y}) = 0, \\ 1 & \text{o.w.} \end{cases}$$

Following Lemma EC.1, for any index i such that $y'_i = 0$, we have $\mathbb{P}(Y_i(t) = 0 | \mathbf{Y}(t-1) = \mathbf{y}) > 0$. For indices i such that $y'_i = 1$, we have $\mathbb{P}(Y_i(t) = 1 | \mathbf{Y}(t-1) = \mathbf{y}) > 0$. Consequently, for all $i \in V$ and $\mathbf{y} \in \{0, 1\}^{|V|}$, we have $\mathbb{P}(Y_i(t) = y'_i | \mathbf{Y}(t-1) = \mathbf{y}) > 0$.

For any arbitrary $\mathbf{y} \in \{0, 1\}^{|V|}$, the transition probability between \mathbf{y} and \mathbf{y}' can be written as:

$$P(\mathbf{y}, \mathbf{y}') = \prod_{i \in V} \mathbb{P}(Y_i(t) = y'_i | \mathbf{Y}(t-1) = \mathbf{y}),$$

which, based on our analysis, will be positive. This implies that all states in the MC communicate with the state \mathbf{y}' . Therefore, the set of states that \mathbf{y}' communicates with forms a recurrent communication class, while all other states are in transient classes.

Furthermore, since $P(\mathbf{y}', \mathbf{y}') > 0$ also holds, state \mathbf{y}' has period 1 which implies that the recurrent communication class of the MC is aperiodic.

In conclusion, this MC has a limiting distribution $\boldsymbol{\pi}$ that satisfies $\boldsymbol{\pi} = \boldsymbol{\pi}P$ and the limiting adoption probability of each agent is a linear transformation of $\boldsymbol{\pi}$ that follows (2). \square

A.2. Extensions to Non-i.i.d. Noise Assumption

In our basic model, we assume that random noise $\epsilon_i(t)$ is i.i.d. for all $i \in V$ and $t \geq 1$. In this section, we explore extensions where this assumption is relaxed. The FPA performance in the absence of i.i.d. assumption depends on the specific definition of the joint distribution. We outline and discuss two cases: The first case considers the situation where the noise is not identically distributed across time, and the second case examines the situation where the noise is not independently distributed across agents at the same time.

A.2.1. Non-identical Noise Assumption. We assume that random noise $\epsilon_i(t)$ for all $i \in V$ and $t \geq 1$ is independently drawn from distributions with varying parameters, but converges to a specific distribution as $t \rightarrow \infty$. Let $F_{\epsilon(t)}(\cdot)$ denote the CDF of random noise at time t , and let $F_\epsilon(\cdot)$ denote the CDF of the asymptotic distribution as $t \rightarrow \infty$. For instance, if we assume that $\epsilon_i(t)$ follows a logistic distribution, we can model it as $\epsilon_i(t) \sim \text{logistic}(0, \sigma^2(t))$ for each agent $i \in V$ at time step $t \geq 1$, where $\sigma(t)$ diminishes over time and eventually converges to a constant σ as $t \rightarrow \infty$.

In this setting, the diffusion process can be viewed as a non-stationary Markov chain, where the transition matrix evolves over time. However, since the transition matrix converges to the one constructed using the asymptotic distribution, the system will eventually behave as though the noise distribution is fixed at the asymptotic distribution, leading to a stable long-term behavior.

As a result, the FPA scheme can easily be generalized from the i.i.d. case by changing the parameter L to denote the maximum Lipschitz constant of all CDFs $F_{\epsilon(t)}(\cdot)$ for $t \geq 1$. Specifically, as long as Assumption 1 is satisfied, we can still construct the FPA solution as in (5), and the approximation error can be bounded as shown in Theorem 1.

Note that requiring L to be the maximum Lipschitz constant is not a stringent condition. For example, in the case of the logistic distribution with decreasing variance, L just corresponds to the Lipschitz constant of the asymptotic distribution $\text{logistic}(0, \sigma^2)$. Therefore, the FPA scheme and the general approximation error analysis remain the same as in the original i.i.d. case. We also want to make the following remark: As

explained in Section 2.1 following Assumption 1, the bound we provide is instance-independent, which means that it assumes that the intrinsic value \mathbf{v} can take arbitrary values. If we consider an instance-dependent bound for a specific \mathbf{v} , the parameter L should be taken as the maximum Lipschitz constant only within the domain $[-\max_{i \in V} v_i - \beta, -\min_{i \in V} v_i]$ of all CDFs. In this case, L may not necessarily correspond to the Lipschitz constant of the asymptotic distribution. As a result, introducing nonstationarity (e.g., time-varying noise) could increase the error bound compared to the stationary case.

A.2.2. Non-independent Noise Assumption. When random noise $\epsilon_i(t)$ is correlated across the network, the problem becomes significantly more complex. The primary challenge is to derive a counterpart to Lemma 2 that accounts for the correlation in the random noise and provides an upper bound for the variance of the local network effect.

We start by defining the correlation across the random noise. Due to the nonlinear transitions and correlated adoption states, it is insufficient to specify just the covariance matrix of the random noise. We must also capture a more specific structure of this correlation. To do so, we introduce a quadrant function $H(x, y)$, which quantifies the joint probability of $\epsilon_i(t)$ and $\epsilon_j(t)$ as follows:

$$H(x, y) := \mathbb{P}(\epsilon_i(t) \leq x, \epsilon_j(t) \leq y) - \mathbb{P}(\epsilon_i(t) \leq x)\mathbb{P}(\epsilon_j(t) \leq y).$$

To quantify the magnitude of the correlation, we introduce a parameter a that represents the uniform upper bound of $H(x, y)$, i.e., for all $x, y \in \mathbb{R}$, we require $0 \leq H(x, y) \leq a$. Notably, a is naturally bounded by $1/4$, since:

$$\begin{aligned} H(x, y) &= \mathbb{P}(\epsilon_i(t) \leq x, \epsilon_j(t) \leq y) - \mathbb{P}(\epsilon_i(t) \leq x)\mathbb{P}(\epsilon_j(t) \leq y) = \text{Cov}(\mathbb{1}\{\epsilon_i(t) \leq x\}, \mathbb{1}\{\epsilon_j(t) \leq y\}) \\ &\leq \sqrt{\text{Var}(\mathbb{1}\{\epsilon_i(t) \leq x\}) \text{Var}(\mathbb{1}\{\epsilon_j(t) \leq y\})} \\ &\leq \frac{1}{4}, \end{aligned}$$

where the first inequality follows from the Cauchy-Schwarz inequality, and the second inequality holds because the variance of a binary random variable is upper bounded by $1/4$.

Next, we present Lemma EC.2, which provides an upper bound for the variance of the local network effect when random noise may be correlated in a single time step.

LEMMA EC.2 (Local Network Effect Variance). *Under Assumptions 1, 2, and $0 \leq H(\cdot, \cdot) \leq a \leq 1/4$, for any diffusion instance $(G, \mathbf{v}, F_\epsilon(\cdot), \beta)$ and $t \geq 1$, the variance of network effect can be upper bounded by*

$$\kappa(t) \leq \left(\frac{1}{4} - a\right) \mathcal{B}(G; \rho) + \frac{a}{1 - \rho^2} \mathbf{e} \leq \left(\frac{1}{4} - a\right) \left(\mathbf{I} - \rho^2 \hat{\mathbf{A}}\right)^{-1} \mathbf{b} + \frac{a}{1 - \rho^2} \mathbf{e}.$$

We observe that when the random noise is correlated, the upper bound for the variance of the local network effect increases, becoming a linear combination of the original bound (under the independent noise assumption) and a vector of ones. When $a = 0$, corresponding to the independent noise case, the result of Lemma EC.2 reduces to the original result in Lemma 2. Using this result, we can derive the associated approximation error for the FPA scheme. Since the procedure follows similarly to the independent case, we omit a detailed demonstration here.

Proof of Lemma EC.2: While much of the proof follows the same structure as in Lemmas 1 and 2, the key difference arises in the decomposition of the covariance in (EC.1).

When random noise is correlated, the first term in (EC.1) will no longer be zero, but it can be upper bounded as follows:

$$\begin{aligned} & \mathbb{E}_{\mathbf{Y}(t-1)} \left[\text{Cov}_{\epsilon(t)} (\mathbb{E}[Y_i(t)|\mathbf{Y}(t-1), \boldsymbol{\epsilon}(t)], \mathbb{E}[Y_{i'}(t)|\mathbf{Y}(t-1), \boldsymbol{\epsilon}(t)]) \middle| \mathbf{Y}(t-1) \right] \\ &= \mathbb{E}_{\mathbf{Y}(t-1)} \left[\text{Cov}_{\epsilon(t)} \left(\mathbb{1} \left\{ \epsilon_i(t) \leq -v_i - \beta \frac{\sum_{j \in \mathcal{N}_i} Y_j(t-1)}{d_i} \right\}, \mathbb{1} \left\{ \epsilon_{i'}(t) \leq -v_{i'} - \beta \frac{\sum_{j' \in \mathcal{N}_{i'}} Y_{j'}(t-1)}{d_{i'}} \right\} \right) \middle| \mathbf{Y}(t-1) \right] \\ &= \mathbb{E}_{\mathbf{Y}(t-1)} \left[H \left(-v_i - \beta \frac{\sum_{j \in \mathcal{N}_i} Y_j(t-1)}{d_i}, -v_{i'} - \beta \frac{\sum_{j' \in \mathcal{N}_{i'}} Y_{j'}(t-1)}{d_{i'}} \right) \middle| \mathbf{Y}(t-1) \right] \\ &\leq a. \end{aligned}$$

Thus, the covariance matrix of the adoption states is upper bounded by

$$\begin{aligned} \boldsymbol{\Sigma}(t) &\leq \rho^2 \tilde{\mathbf{A}} \boldsymbol{\Sigma}(t-1) \tilde{\mathbf{A}}^\top + \left(\frac{1}{4} - a \right) \mathbf{I} + a \mathbf{e} \mathbf{e}^\top \\ &\leq \rho^4 \tilde{\mathbf{A}}^2 \boldsymbol{\Sigma}(t-2) (\tilde{\mathbf{A}}^\top)^2 + \left(\frac{1}{4} - a \right) (\mathbf{I} + \rho^2 \tilde{\mathbf{A}} \tilde{\mathbf{A}}^\top) + a(1 + \rho^2) \mathbf{e} \mathbf{e}^\top \leq \dots \\ &\leq \left(\frac{1}{4} - a \right) \left(\mathbf{I} + \sum_{\tau=1}^{t-1} \rho^{2\tau} \tilde{\mathbf{A}}^\tau (\tilde{\mathbf{A}}^\top)^\tau \right) + a \left(1 + \sum_{\tau=1}^{t-1} \rho^{2\tau} \right) \mathbf{e} \mathbf{e}^\top. \end{aligned}$$

Consequently, the variance of the local network effect can be upper bounded by

$$\begin{aligned} \boldsymbol{\kappa}(t) &\leq \left(\frac{1}{4} - a \right) \left[\sum_{\tau=1}^t \rho^{2\tau-2} \tilde{\mathbf{A}}^\tau \odot (\tilde{\mathbf{A}}^\top)^\tau \right] + a \sum_{\tau=0}^{t-1} \rho^{2\tau} \mathbf{e} \\ &\leq \left(\frac{1}{4} - a \right) \left[\mathbf{I} + \sum_{\tau=1}^{t-1} \rho^{2\tau} \tilde{\mathbf{A}}^\tau \right] \mathbf{b} + a \sum_{\tau=0}^{t-1} \rho^{2\tau} \mathbf{e}. \end{aligned}$$

In conclusion, we have

$$\boldsymbol{\kappa}(t) \leq \left(\frac{1}{4} - a \right) \mathcal{B}(G; \rho) + \frac{a}{1 - \rho^2} \mathbf{e} \leq \left(\frac{1}{4} - a \right) (\mathbf{I} - \rho^2 \tilde{\mathbf{A}})^{-1} \mathbf{b} + \frac{a}{1 - \rho^2} \mathbf{e}.$$

□

Appendix B: Supporting Arguments for Section 3

In this Section, we provide the proofs for the theoretical results in Section 3. The main objective of Section 3 is to prove our key result, Theorem 1. Due to the length of the proof, we decompose it into three parts for clarity and ease of reading. We begin by establishing the existence and uniqueness of the FPA solution in Appendix B.1. Next, we analyze the covariance across agents, which is the most complex part, in Appendix B.2. Finally, we conclude the proof for Theorem 1 and its associated corollary in Appendix B.3.

B.1. Supporting Arguments for Proposition 2

Proof of Proposition 2: We first show the property (i) and then prove properties (ii) and (iii) by showing that $\mathbf{h}(\cdot)$ is a contraction mapping.

Proof of (i): When $\mathbf{a} \leq \mathbf{b}$, we have $\sum_{j \in \mathcal{N}_i} a_j \leq \sum_{j \in \mathcal{N}_i} b_j$ for all $i \in V$. Since CDF $F_\epsilon(\cdot)$ is monotonically nondecreasing, if $\mathbf{a} \leq \mathbf{b}$,

$$1 - F_\epsilon \left(-v_i - \beta \frac{\sum_{j \in \mathcal{N}_i} a_j}{d_i} \right) \leq 1 - F_\epsilon \left(-v_i - \beta \frac{\sum_{j \in \mathcal{N}_i} b_j}{d_i} \right),$$

for all $i \in V$, which implies $\mathbf{h}(\mathbf{a}) \leq \mathbf{h}(\mathbf{b})$.

Proof of (ii) and (iii): It is trivial that $\mathbf{h}(\cdot)$ maps \mathbb{R}^n to itself. Consider the Jacobian matrix of $\mathbf{h}(\boldsymbol{\mu})$, for an arbitrary $\boldsymbol{\mu} \in \mathbb{R}^n$, we have

$$\frac{\partial h(\boldsymbol{\mu})_i}{\partial \mu_j} = \begin{cases} 0, & j \notin \mathcal{N}_i \\ \frac{\beta}{d_i} \frac{\partial F_\epsilon \left(-v_i - \beta \frac{\sum_{j' \in \mathcal{N}_i} \mu_{j'}}{d_i} \right)}{\partial \left(-v_i - \beta \frac{\sum_{j' \in \mathcal{N}_i} \mu_{j'}}{d_i} \right)}, & j \in \mathcal{N}_i. \end{cases}$$

By Assumption 1, we can have $\left| \frac{\partial h(\boldsymbol{\mu})_i}{\partial \mu_j} \right| \leq \frac{\beta L}{d_i}$ for all $j \in \mathcal{N}_i$. Therefore, the ℓ_∞ -norm of $\frac{d\mathbf{h}(\boldsymbol{\mu})}{d\boldsymbol{\mu}}$ can be upper bounded as

$$\left\| \frac{d\mathbf{h}(\boldsymbol{\mu})}{d\boldsymbol{\mu}} \right\|_\infty = \max_{i \in V} \sum_{j \in V} \left| \frac{\partial h(\boldsymbol{\mu})_i}{\partial \mu_j} \right| \leq \max_{i \in V} d_i \frac{\beta L}{d_i} = \beta L < 1$$

where the last inequality follows from Assumption 2.

Thus, for all $\boldsymbol{\mu} \in \mathbb{R}^n$, we have $\left\| \frac{d\mathbf{h}(\boldsymbol{\mu})}{d\boldsymbol{\mu}} \right\|_\infty < 1$. It then implies that $\mathbf{h}(\boldsymbol{\mu})$ is a contraction mapping. By the contraction mapping theorem, we conclude the proof. \square

B.2. Supporting Arguments for Covariance Analysis

Before we present the proof for Lemma 1, we would like to first present a series of auxiliary definitions and lemmas. To begin with, we define the concept of positive quadrant dependent and positive association in Definitions EC.1 and EC.2, respectively.

DEFINITION EC.1 (POSITIVE QUADRANT DEPENDENT, NEWMAN (1980)). A pair of random variables (X, Y) is positively quadrant dependent (PQD) if

$$\mathbb{P}(X \leq a, Y \leq b) - \mathbb{P}(X \leq a)\mathbb{P}(Y \leq b) \geq 0, \quad \forall a, b \in \mathbb{R}.$$

DEFINITION EC.2 (POSITIVE ASSOCIATION, ESARY ET AL. (1967)). A random vector $\mathbf{X} = (X_1, \dots, X_n) \in \mathbb{R}^n$ is said to be positively associated (PA) whenever

$$\text{Cov}(f(\mathbf{X}), g(\mathbf{X})) \geq 0,$$

for all coordinate-wise nondecreasing functions f and g on \mathbb{R}^n such that $f(\mathbf{X})$ and $g(\mathbf{X})$ both possess finite second moments.

We then present several lemmas on the properties of variables that are PQD and PA, which paves the way for our covariance analysis.

LEMMA EC.3 (ESARY ET AL. (1967), P2). *If two sets of PA random variables are independent of one another, then their union is a set of PA random variables.*

Proof of Lemma EC.3: Let $\mathbf{X} = (X_1, X_2, \dots, X_n)$ and $\mathbf{Y} = (Y_1, Y_2, \dots, Y_m)$ be PA random vectors, and suppose that \mathbf{X} and \mathbf{Y} are independent of each other. Let f, g be any piecewise nondecreasing functions with regard to (\mathbf{X}, \mathbf{Y}) . Using the law of total covariance, we can decompose the covariance as follows:

$$\text{Cov}(f(\mathbf{X}, \mathbf{Y}), g(\mathbf{X}, \mathbf{Y})) = \mathbb{E}_{\mathbf{X}} [\text{Cov}_{\mathbf{Y}}(f(\mathbf{X}, \mathbf{Y}), g(\mathbf{X}, \mathbf{Y}) | \mathbf{X})] + \text{Cov}_{\mathbf{X}} (\mathbb{E}_{\mathbf{Y}}[f(\mathbf{X}, \mathbf{Y}) | \mathbf{X}], \mathbb{E}_{\mathbf{Y}}[g(\mathbf{X}, \mathbf{Y}) | \mathbf{X}]).$$

Then, it is straightforward to verify that both terms are nonnegative.

(i) First term: Since \mathbf{Y} is PA, we have $\text{Cov}_{\mathbf{Y}}(f(\mathbf{x}, \mathbf{Y}), g(\mathbf{x}, \mathbf{Y})) \geq 0$ for any given \mathbf{x} . Thus, by taking the expectation over this covariance, the first term is nonnegative.

(ii) Second term: Since \mathbf{X} and \mathbf{Y} are independent, thus, we have

$$\mathbb{E}_{\mathbf{Y}}[f(\mathbf{X}, \mathbf{Y}) | \mathbf{X} = \mathbf{x}] = \mathbb{E}_{\mathbf{Y}}[f(\mathbf{x}, \mathbf{Y})].$$

For any arbitrary $\mathbf{x} \in \mathbb{R}^n$, $\mathbb{E}_{\mathbf{Y}}[f(\mathbf{x}, \mathbf{Y})]$ and $\mathbb{E}_{\mathbf{Y}}[g(\mathbf{x}, \mathbf{Y})]$ are nondecreasing functions with regard to \mathbf{x} . Since \mathbf{X} is PA random vector, the covariance between these two nondecreasing functions will also be nonnegative.

In conclusion, the union set of these random variables, i.e., (\mathbf{X}, \mathbf{Y}) , is PA. \square

LEMMA EC.4 (Goldstein and Wiroonsri (2018), Lemma 3.3). *The pair of random variables $(f(\mathbf{X}), g(\mathbf{X}))$ is PQD whenever random vector \mathbf{X} is PA, and $f(\cdot)$ and $g(\cdot)$ are coordinate-wise nondecreasing functions of \mathbf{X} .*

Proof of Lemma EC.4: In order to prove that $(f(\mathbf{X}), g(\mathbf{X}))$ is PQD, we need to show that

$$\mathbb{P}(f(\mathbf{X}) \leq a, g(\mathbf{X}) \leq b) - \mathbb{P}(f(\mathbf{X}) \leq a)\mathbb{P}(g(\mathbf{X}) \leq b) \geq 0, \quad \forall a, b \in \mathbb{R}.$$

For any given pair of $(a, b) \in \mathbb{R}^2$, let's define the following nondecreasing functions:

$$f_a(x) = \begin{cases} 0 & x \leq a, \\ 1 & x > a, \end{cases} \text{ and } g_b(x) = \begin{cases} 0 & x \leq b, \\ 1 & x > b. \end{cases}$$

As a consequence, the composite functions $f_a(f(\cdot))$ and $g_b(g(\cdot))$ are also both nondecreasing functions.

Since the random vector \mathbf{X} is PA, we have

$$\begin{aligned} 0 &\leq \text{Cov}(f_a(f(\mathbf{X})), g_b(g(\mathbf{X}))) = \text{Cov}\left(1 - f_a(f(\mathbf{X})), 1 - g_b(g(\mathbf{X}))\right) \\ &= \mathbb{P}\left(1 - f_a(f(\mathbf{X})) = 1, 1 - g_b(g(\mathbf{X})) = 1\right) - \mathbb{P}\left(1 - f_a(f(\mathbf{X})) = 1\right)\mathbb{P}\left(1 - g_b(g(\mathbf{X})) = 1\right) \\ &= \mathbb{P}\left(f(\mathbf{X}) \leq a, g(\mathbf{X}) \leq b\right) - \mathbb{P}\left(f(\mathbf{X}) \leq a\right)\mathbb{P}\left(g(\mathbf{X}) \leq b\right), \end{aligned}$$

where the first inequality follows from the definition of PA random variables and the remaining equalities follow from the definition of covariance.

As a consequence, we have proved that $(f(\mathbf{X}), g(\mathbf{X}))$ is PQD. \square

LEMMA EC.5 (Goldstein and Wiroonsri (2018), Lemma 3.1; Newman (1980), Lemma 3). *Let the random variables X, Y have finite second moments and be PQD. Then for any real-valued functions $f(\cdot)$ and $g(\cdot)$ on \mathbb{R} with their derivatives $f'(\cdot), g'(\cdot)$ bounded,*

$$|\text{Cov}(f(X), g(Y))| \leq \|f'\|_{\infty} \|g'\|_{\infty} \text{Cov}(X, Y),$$

where $\|\cdot\|_{\infty}$ denotes the sup norm on \mathbb{R} .

Proof of Lemma EC.5: Let $H_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y) - \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y)$.

The covariance of $f(X)$ and $g(Y)$ can be written as

$$\text{Cov}(f(X), g(Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H_{X,Y}(x, y) f'(x) g'(y) dx dy.$$

Thus, by the positivity of $H_{X,Y}$ for PQD random variables, we can arrive at

$$|\text{Cov}(f(X), g(Y))| \leq \|f'\|_{\infty} \|g'\|_{\infty} \cdot \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H_{X,Y}(x, y) dx dy = \|f'\|_{\infty} \|g'\|_{\infty} \text{Cov}(X, Y).$$

\square

COROLLARY EC.1 (Lower Bound of Covariance). *Let the random variables $X \in \mathcal{X} \subseteq \mathbb{R}$, $Y \in \mathcal{Y} \subseteq \mathbb{R}$ be bounded and be PQD. Then for any real-valued functions $f(\cdot) : \mathcal{X} \rightarrow \mathbb{R}$ and $g(\cdot) : \mathcal{Y} \rightarrow \mathbb{R}$ with their derivatives $f'(\cdot) \geq \lambda_f \geq 0$, $g'(\cdot) \geq \lambda_g \geq 0$ and bounded,*

$$\text{Cov}(f(X), g(Y)) \geq \lambda_f \lambda_g \text{Cov}(X, Y).$$

Proof of Corollary EC.1: Let $H_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y) - \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y)$.

The covariance of $f(X)$ and $g(Y)$ can be written as

$$\text{Cov}(f(X), g(Y)) = \int_{x \in \mathcal{X}} \int_{y \in \mathcal{Y}} H_{X,Y}(x, y) f'(x) g'(y) dx dy.$$

Thus, by the positivity of $H_{X,Y}$ for PQD random variables, we can arrive at

$$\text{Cov}(f(X), g(Y)) \geq \lambda_f \lambda_g \cdot \int_{x \in \mathcal{X}} \int_{y \in \mathcal{Y}} H_{X,Y}(x, y) dx dy = \lambda_f \lambda_g \text{Cov}(X, Y).$$

□

The proof of Corollary EC.1 is similar to that of Lemma EC.5, mainly relying on the definition of PQD random variables.

Equipped with these properties, we then prove the upper bound for the covariance matrix in Lemma 1.

Proof of Lemma 1: This proof of the upper bound of the covariance matrix involves two critical steps. First, we need to establish that the adoption state vector $\mathbf{Y}(t) = (Y_1(t), Y_2(t), \dots, Y_n(t))$ is positively associated for any time $t \geq 0$. Second, we iteratively bound the covariance matrix at each time step, leveraging the PA property.

Positively associated variables: We will show this by induction.

Base case: $t = 0$. The random vector $\mathbf{Y}(0)$ is independent, so it is PA.

To show time $t = s + 1$: Assume that the random vector $\mathbf{Y}(s)$ is PA.

The random noise $\epsilon_i(s+1)$ is independent across all agents $i \in V$ and therefore, the vector $\epsilon(s+1)$ is PA. Furthermore, vectors $\mathbf{Y}(s)$ and $\epsilon(s+1)$ are independent of each other. By applying Lemma EC.3, we can show that the union set of random variables $\mathbf{Y}(s)$ and $\epsilon(s+1)$ is also PA.

For any $i \in V$, the random variable

$$Y_i(s+1) = \mathbb{1} \left\{ v_i + \beta \frac{\sum_{j \in \mathcal{N}_i} Y_j(s)}{d_i} + \epsilon_i(s+1) \geq 0 \right\},$$

can be considered as a coordinate-wise nondecreasing function of both $\mathbf{Y}(s)$ and $\epsilon(s+1)$. Let $\mathbf{Y}(s+1) := \phi(\mathbf{Y}(s), \epsilon(s+1)) \in \mathbb{R}^N$ denote this transformation. It can be easily shown that the random vector $\mathbf{Y}(s+1)$ is also a PA vector. Specifically, for any nondecreasing functions f and g ,

$$\text{Cov}(f(\mathbf{Y}(s+1)), g(\mathbf{Y}(s+1))) = \text{Cov} \left(f(\phi(\mathbf{Y}(s), \epsilon(s+1))), g(\phi(\mathbf{Y}(s), \epsilon(s+1))) \right) \geq 0,$$

where the inequality follows from the definition of PA random variables. Since the composite functions $f(\phi(\cdot))$ and $g(\phi(\cdot))$ are coordinate-wise nondecreasing and the union of $\mathbf{Y}(s)$ and $\epsilon(s+1)$ is PA, the covariance is nonnegative.

As a result, we have shown that $\mathbf{Y}(t)$ is PA for all $t \geq 0$. With this result, we can apply Lemma EC.4 to demonstrate that the network effect terms $\sum_{j \in \mathcal{N}_i} Y_j(t)$ and $\sum_{j' \in \mathcal{N}_{i'}} Y_{j'}(t)$ are PQD for all $i, i' \in V$ and $t \geq 0$.

Upper bound of the covariance matrix:

In the following, we will iteratively upper bound the covariance between any pair of $(Y_i(t), Y_{i'}(t))$ for all $i, i' \in V$ and $t \geq 0$.

Using the law of total covariance, we can decompose $\text{Cov}(Y_i(t), Y_{i'}(t))$ into two parts:

$$\begin{aligned} \text{Cov}(Y_i(t), Y_{i'}(t)) &= \mathbb{E}_{\mathbf{Y}(t-1)} \left[\text{Cov}_{\epsilon(t)} (Y_i(t), Y_{i'}(t) \mid \mathbf{Y}(t-1)) \right] \\ &\quad + \text{Cov}_{\mathbf{Y}(t-1)} \left(\mathbb{E}_{\epsilon(t)} [Y_i(t) \mid \mathbf{Y}(t-1)], \mathbb{E}_{\epsilon(t)} [Y_{i'}(t) \mid \mathbf{Y}(t-1)] \right). \end{aligned}$$

The first term $\mathbb{E}_{\mathbf{Y}(t-1)} [\text{Cov}_{\epsilon(t)} (Y_i(t), Y_{i'}(t) \mid \mathbf{Y}(t-1))]$ is always 0. The reason is as follows: by applying the law of total conditional covariance again, we have

$$\begin{aligned} &\text{Cov}_{\epsilon(t)} (Y_i(t), Y_{i'}(t) \mid \mathbf{Y}(t-1)) \\ &= \mathbb{E}_{\epsilon(t)} \left[\text{Cov} (Y_i(t), Y_{i'}(t) \mid \mathbf{Y}(t-1), \epsilon(t)) \mid \mathbf{Y}(t-1) \right] \\ &\quad + \text{Cov}_{\epsilon(t)} \left(\mathbb{E}[Y_i(t) \mid \mathbf{Y}(t-1), \epsilon(t)], \mathbb{E}[Y_{i'}(t) \mid \mathbf{Y}(t-1), \epsilon(t)] \mid \mathbf{Y}(t-1) \right). \end{aligned} \tag{EC.1}$$

The former term vanishes because $Y_i(t)$ and $Y_{i'}(t)$ are deterministic when given $\mathbf{Y}(t-1), \epsilon_t$. The latter term is also zero since $\epsilon_i(t)$ and $\epsilon_{i'}(t)$ are independent of each other.

We then show that the second term $\text{Cov}_{\mathbf{Y}(t-1)} \left(\mathbb{E}_{\epsilon(t)} [Y_i(t) \mid \mathbf{Y}(t-1)], \mathbb{E}_{\epsilon(t)} [Y_{i'}(t) \mid \mathbf{Y}(t-1)] \right)$ can be bounded recursively as

$$\begin{aligned} &\text{Cov}_{\mathbf{Y}(t-1)} \left(\mathbb{E}_{\epsilon(t)} [Y_i(t) \mid \mathbf{Y}(t-1)], \mathbb{E}_{\epsilon(t)} [Y_{i'}(t) \mid \mathbf{Y}(t-1)] \right) \\ &= \text{Cov} \left(1 - F_\epsilon \left(-v_i - \beta \frac{\sum_{j \in \mathcal{N}_i} Y_j(t-1)}{d_i} \right), 1 - F_\epsilon \left(-v_{i'} - \beta \frac{\sum_{j' \in \mathcal{N}_{i'}} Y_{j'}(t-1)}{d_{i'}} \right) \right) \\ &\leq \frac{(L\beta)^2}{d_i d_{i'}} \text{Cov} \left(\sum_{j \in \mathcal{N}_i} Y_j(t-1), \sum_{j' \in \mathcal{N}_{i'}} Y_{j'}(t-1) \right) \end{aligned} \tag{EC.2a}$$

$$= \frac{\rho^2}{d_i d_{i'}} \sum_{j \in \mathcal{N}_i} \sum_{j' \in \mathcal{N}_{i'}} \text{Cov} (Y_j(t-1), Y_{j'}(t-1)). \tag{EC.2b}$$

where (EC.2a) follows from Lemma EC.5 and (EC.2b) follows from the definition of covariance of linear combinations and the fact that $\rho = L\beta$.

So far, we have proved that the covariance of $Y_i(t)$ and $Y_{i'}(t)$ for any pair (i, i') where $i, i' \in V$ and $i \neq i'$ can be upper bounded as follows:

$$\text{Cov}(Y_i(t), Y_{i'}(t)) \leq \frac{\rho^2}{d_i d_{i'}} \sum_{j \in \mathcal{N}_i} \sum_{j' \in \mathcal{N}_{i'}} \text{Cov} (Y_j(t-1), Y_{j'}(t-1)). \tag{EC.3}$$

Meanwhile, for any $i \in V$, it is trivial that $\text{Var}(Y_i(t))$ can be upper bounded by constant $1/4$ as $Y_i(t)$ is a binary variable.

When $t = 1$, the covariance matrix can be upper bounded by $1/4 \cdot \mathbf{I}$ as all the agents' decisions are not correlated. Recalling the definition of $\tilde{\mathbf{A}}$, iteratively, we can upper bound the covariance matrix $\Sigma(t)$ as

$$\Sigma(t) \leq \frac{1}{4} \mathbf{I} + \rho^2 \tilde{\mathbf{A}} \Sigma(t-1) \tilde{\mathbf{A}}^\top \leq \frac{1}{4} \mathbf{I} + \rho^2 \tilde{\mathbf{A}} \left[\frac{1}{4} \mathbf{I} + \rho^2 \tilde{\mathbf{A}} \Sigma(t-2) \tilde{\mathbf{A}}^\top \right] \tilde{\mathbf{A}}^\top \leq \dots \leq \frac{1}{4} \left[\mathbf{I} + \sum_{\tau=1}^{t-1} \rho^{2\tau} \tilde{\mathbf{A}}^\top (\tilde{\mathbf{A}}^\top)^\tau \right],$$

where the inequalities follow since (EC.3) and the trivial $1/4$ bound for the variance of binary variables. \square

Proof of Lemma 2: Based on Lemma 1, we can upper bound $\kappa_i(t)$ by the i -th diagonal entry of matrix

$$\tilde{\mathbf{A}}\Sigma(t)\tilde{\mathbf{A}}^\top = \frac{1}{4} \left[\sum_{\tau=1}^t \rho^{2\tau-2} \tilde{\mathbf{A}}^\tau \left(\tilde{\mathbf{A}}^\top \right)^\tau \right].$$

Hereafter, we first reformulate a closed-form expression for this upper bound and provide a relaxation for this upper bound that can be calculated much easier.

Let \mathbf{e}_i denote the vector when only the i -th entry equals 1 and other entries are 0. For an arbitrary matrix \mathbf{M} , the i -th diagonal entry of matrix $\mathbf{M}\mathbf{M}^\top$ equals to

$$\mathbf{e}_i^\top \mathbf{M}\mathbf{M}^\top \mathbf{e}_i = \sum_{j=1}^n M_{ij}^2.$$

Therefore, we can represent the vector of all diagonal entries to be $(\mathbf{M} \odot \mathbf{M})\mathbf{e}$, where \odot means Hadamard product. As a consequence, we arrive at

$$\boldsymbol{\kappa}(t) \leq \frac{1}{4} \left[\sum_{\tau=1}^t \rho^{2\tau-2} \tilde{\mathbf{A}}^\tau \odot \tilde{\mathbf{A}}^\tau \right] \mathbf{e}.$$

This upper bound involves the sum of a series of vectors, which cannot be reduced to a simplified expression. For a better interpretation of our results, we further relate it to Neumann series, so that it can be represented by a much compact form. Note that, for all $i, j \in V$,

$$\left(\tilde{\mathbf{A}}\tilde{\mathbf{A}}^\top \right)_{ij} = \frac{1}{d_i d_j} \sum_{k=1}^n \mathbb{1}\{k \text{ is the common in-neighbors of } i \text{ and } j\} \leq \frac{1}{d_i},$$

where the inequality follows since the number of common in-neighbors of agents i and j is smaller than d_j . As a result, $\tilde{\mathbf{A}}\tilde{\mathbf{A}}^\top \leq \mathbf{b}\mathbf{e}^\top$. Furthermore, for all $t \geq 1$, we can show that,

$$\tilde{\mathbf{A}} \left(\tilde{\mathbf{A}}^\top \right)^t = \tilde{\mathbf{A}}\tilde{\mathbf{A}}^\top \left(\tilde{\mathbf{A}}^\top \right)^{t-1} \leq \mathbf{b}\mathbf{e}^\top \left(\tilde{\mathbf{A}}^\top \right)^{t-1} = \mathbf{b} \left(\tilde{\mathbf{A}}^{t-1} \mathbf{e} \right)^\top = \mathbf{b}\mathbf{e}^\top,$$

where the inequality holds since both matrices $\mathbf{b}\mathbf{e}^\top$ and $\tilde{\mathbf{A}}$ have all entries to be nonnegative, the last equality holds since $\tilde{\mathbf{A}}$ is a row-stochastic matrix.

Consequently, for all $i \in V$, we have

$$\mathbf{e}_i^\top \tilde{\mathbf{A}}^\tau \left(\tilde{\mathbf{A}}^\top \right)^\tau \mathbf{e}_i \leq \mathbf{e}_i^\top \tilde{\mathbf{A}}^{\tau-1} \mathbf{b} \mathbf{e}^\top \mathbf{e}_i = \mathbf{e}_i^\top \tilde{\mathbf{A}}^{\tau-1} \mathbf{b}, \quad (\text{EC.4})$$

where the inequality follows since all entries of $\tilde{\mathbf{A}}$ and \mathbf{e}_i are nonnegative.

In conclusion, we can derive the relaxed upper bound for $\boldsymbol{\kappa}(t)$ as

$$\boldsymbol{\kappa}(t) \leq \frac{1}{4} \left[\sum_{\tau=1}^t \rho^{2\tau-2} \tilde{\mathbf{A}}^\tau \odot \left(\tilde{\mathbf{A}}^\top \right)^\tau \right] \mathbf{e} \leq \frac{1}{4} \left[\sum_{\tau=1}^t \rho^{2\tau-2} \tilde{\mathbf{A}}^{\tau-1} \right] \mathbf{b} = \frac{1}{4} \left[\mathbf{I} + \sum_{\tau=1}^{t-1} \rho^{2\tau} \tilde{\mathbf{A}}^\tau \right] \mathbf{b},$$

where the second inequality follows from (EC.4).

□

B.3. Supporting Arguments for Theorem 1

Proof of Lemma 3: Let $\Delta_i(t) = \frac{\beta}{d_i}(\sum_{j \in \mathcal{N}_i} q_j(t-1) - \sum_{j \in \mathcal{N}_i} Y_j(t-1))$. For any $i \in V$ and $t \geq 0$, the adoption probability of agent i at time t can be written as

$$\begin{aligned} q_i(t) &= \mathbb{E}_{\mathbf{Y}(t-1)} \left[\mathbb{E}_{\epsilon(t)} [Y_i(t) | \mathbf{Y}(t-1)] \right] = \mathbb{E}_{\mathbf{Y}(t-1)} \left[1 - F_\epsilon \left(-v_i - \beta \frac{\sum_{j \in \mathcal{N}_i} Y_j(t-1)}{d_i} \right) \right] \\ &= 1 - \mathbb{E}_{\mathbf{Y}(t-1)} \left[F_\epsilon \left(-v_i - \beta \frac{\sum_{j \in \mathcal{N}_i} q_j(t-1)}{d_i} + \Delta_i(t-1) \right) \right]. \end{aligned}$$

Therefore, we have

$$\begin{aligned} &\left| \mathbb{E}_{\mathbf{Y}(t-1)} \left[F_\epsilon \left(-v_i - \beta \frac{\sum_{j \in \mathcal{N}_i} q_j(t-1)}{d_i} + \Delta_i(t-1) \right) - F_\epsilon \left(-v_i - \beta \frac{\sum_{j \in \mathcal{N}_i} q_j(t-1)}{d_i} \right) \right] \right| \\ &= \sqrt{\left(\mathbb{E}_{\mathbf{Y}(t-1)} \left[F_\epsilon \left(-v_i - \beta \frac{\sum_{j \in \mathcal{N}_i} q_j(t-1)}{d_i} + \Delta_i(t-1) \right) - F_\epsilon \left(-v_i - \beta \frac{\sum_{j \in \mathcal{N}_i} q_j(t-1)}{d_i} \right) \right] \right)^2} \\ &\leq \sqrt{\mathbb{E}_{\mathbf{Y}(t-1)} \left[\left(F_\epsilon \left(-v_i - \beta \frac{\sum_{j \in \mathcal{N}_i} q_j(t-1)}{d_i} + \Delta_i(t-1) \right) - F_\epsilon \left(-v_i - \beta \frac{\sum_{j \in \mathcal{N}_i} q_j(t-1)}{d_i} \right) \right)^2 \right]} \\ &\leq \sqrt{\mathbb{E}_{\mathbf{Y}(t-1)} \left[L^2 |\Delta_i(t-1)|^2 \right]} = \sqrt{\mathbb{E}_{\mathbf{Y}(t-1)} \left[(L\beta)^2 \left(\frac{\sum_{j \in \mathcal{N}_i} q_j(t-1)}{d_i} - \frac{\sum_{j \in \mathcal{N}_i} Y_j(t-1)}{d_i} \right)^2 \right]} \\ &= \sqrt{\rho^2 \text{Var} \left[\frac{\sum_{j \in \mathcal{N}_i} Y_j(t-1)}{d_i} \right]}, \end{aligned}$$

where the first inequality follows by Jensen's inequality and the second inequality follows by Assumption 1.

Since $\delta = \frac{\rho}{2} [\mathcal{B}(G; \rho)]^{\frac{1}{2}}$, by Lemma 2, we can obtain

$$\left| \mathbb{E}_{\mathbf{Y}(t-1)} \left[F_\epsilon \left(-v_i - \beta \frac{\sum_{j \in \mathcal{N}_i} q_j(t-1)}{d_i} + \Delta_i(t-1) \right) - F_\epsilon \left(-v_i - \beta \frac{\sum_{j \in \mathcal{N}_i} q_j(t-1)}{d_i} \right) \right] \right| \leq \delta_i,$$

which further leads to

$$1 - F_\epsilon \left(-v_i - \beta \frac{\sum_{j \in \mathcal{N}_i} q_j(t-1)}{d_i} \right) - \delta_i \leq q_i(t) \leq 1 - F_\epsilon \left(-v_i - \beta \frac{\sum_{j \in \mathcal{N}_i} q_j(t-1)}{d_i} \right) + \delta_i.$$

In summary, we have

$$\left| \mathbf{h}(\mathbf{q}(t-1)) - \mathbf{q}(t) \right|_{\text{ew}} \leq \delta = \frac{\rho}{2} [\mathcal{B}(G; \rho)]^{\frac{1}{2}} \leq \frac{\rho}{2} \left[(\mathbf{I} - \rho^2 \tilde{\mathbf{A}})^{-1} \mathbf{b} \right]_{\text{ew}}^{\frac{1}{2}}.$$

and this concludes the proof. \square

Finally, we prove our main result, i.e., Theorem 1.

Proof of Theorem 1: We first show by induction that, $\underline{\mu}(t) \leq \mathbf{q}(t) \leq \bar{\mu}(t)$ for each $t \geq 0$.

Base case: $t = \mathbf{0}$. By definition, we have $\underline{\mu}(0) = \mathbf{q}(0) = \bar{\mu}(0)$.

To show $t = s + 1$: Assume that $\underline{\mu}(t) \leq \mathbf{q}(t) \leq \bar{\mu}(t)$. Then we have

$$\underline{\mu}(s+1) = \mathbf{h}_{-\delta}(\underline{\mu}(s)) \leq \mathbf{h}_{-\delta}(\mathbf{q}(s)) \leq \mathbf{q}(s+1) \leq \mathbf{h}_\delta(\mathbf{q}(s)) \leq \mathbf{h}_\delta(\bar{\mu}(s)) = \bar{\mu}(s+1),$$

where the first and last inequalities follow from Proposition 2(i) while the other two follow from Lemma 3.

Also, clearly Proposition 2 implies that $\underline{\mu}(t) \leq \mu(t) \leq \bar{\mu}(t)$ for all $t \geq 0$. By the contraction mapping theorem, we know that $\underline{\mu}(t)$ (resp. $\bar{\mu}(t)$) converges to $\underline{\mu}^*$ (resp. $\bar{\mu}^*$) where $\underline{\mu}^*$ (resp. $\bar{\mu}^*$) is the fixed-point solution for $\mathbf{h}_{-\delta}(\underline{\mu}^*) = \underline{\mu}^*$ (resp. $\mathbf{h}_\delta(\bar{\mu}^*) = \bar{\mu}^*$). Thus, the following result holds,

$$\underline{\mu}^* \leq \mathbf{q}^* \leq \bar{\mu}^* \text{ and } \underline{\mu}^* \leq \mu^* \leq \bar{\mu}^*. \quad (\text{EC.5})$$

By the definition of AEOs, the difference between $\underline{\mu}^*$ and $\bar{\mu}^*$ can be written as

$$\bar{\mu}^* - \underline{\mu}^* = \mathbf{h}(\bar{\mu}^*) - \mathbf{h}(\underline{\mu}^*) + 2\delta.$$

Let $\Delta\mu = \bar{\mu}^* - \underline{\mu}^*$. For all $i \in V$,

$$|\Delta\mu_i| \leq \rho \left| \frac{\sum_{j \in \mathcal{N}_i} \bar{\mu}_j^*}{d_i} - \frac{\sum_{j \in \mathcal{N}_i} \underline{\mu}_j^*}{d_i} \right| + 2\delta_i = \rho \left| \frac{\sum_{j \in \mathcal{N}_i} \Delta\mu_j}{d_i} \right| + 2\delta_i, \quad (\text{EC.6})$$

where the inequality comes from Assumption 1.

In matrix form, we can write it as $|\Delta\mu|_{\text{ew}} \leq \rho \tilde{\mathbf{A}} |\Delta\mu|_{\text{ew}} + 2\delta$ or equivalently

$$(\mathbf{I} - \rho \tilde{\mathbf{A}}) |\Delta\mu|_{\text{ew}} \leq 2\delta. \quad (\text{EC.7})$$

Recall that the inverse matrix $(\mathbf{I} - \rho \tilde{\mathbf{A}})^{-1}$ can be expanded into the Neumann series $\mathbf{I} + \sum_{\ell=1}^{\infty} \rho^\ell \tilde{\mathbf{A}}^\ell$. Given that all elements of $\tilde{\mathbf{A}}$ are non-negative, it follows that all elements of $(\mathbf{I} - \rho \tilde{\mathbf{A}})^{-1}$ are also non-negative. Therefore, when we pre-multiply both sides of (EC.7) by $(\mathbf{I} - \rho \tilde{\mathbf{A}})^{-1}$, we obtain the inequality

$$|\Delta\mu|_{\text{ew}} \leq 2 \left(\mathbf{I} - \rho \tilde{\mathbf{A}} \right)^{-1} \delta. \quad (\text{EC.8})$$

Combining (EC.5) and (EC.8), we finally have the following chain of inequalities:

$$\begin{aligned} |\mathbf{q}^* - \mu^*|_{\text{ew}} &\leq |\Delta\mu|_{\text{ew}} \leq \rho (\mathbf{I} - \rho \tilde{\mathbf{A}})^{-1} [\mathcal{B}(G; \rho)]_{\text{ew}}^{\frac{1}{2}} \leq \rho \left(\mathbf{I} - \rho \tilde{\mathbf{A}} \right)^{-1} \left[\left(\mathbf{I} - \rho^2 \tilde{\mathbf{A}} \right)^{-1} \mathbf{b} \right]_{\text{ew}}^{\frac{1}{2}} \\ &= \frac{\rho}{1-\rho} \left[(1-\rho) \left(\mathbf{I} - \rho \tilde{\mathbf{A}} \right)^{-1} \right] \cdot \left[\left(\mathbf{I} - \rho^2 \tilde{\mathbf{A}} \right)^{-1} \mathbf{b} \right]_{\text{ew}}^{\frac{1}{2}} \leq \frac{\rho}{1-\rho} \left[(1-\rho) \left(\mathbf{I} - \rho \tilde{\mathbf{A}} \right)^{-1} \left(\mathbf{I} - \rho^2 \tilde{\mathbf{A}} \right)^{-1} \mathbf{b} \right]_{\text{ew}}^{\frac{1}{2}} \\ &= \frac{\rho}{\sqrt{1-\rho}} \left[\left(\sum_{s=0}^{\infty} \rho^s \tilde{\mathbf{A}}^s \right) \left(\sum_{t=0}^{\infty} \rho^{2t} \tilde{\mathbf{A}}^t \right) \mathbf{b} \right]_{\text{ew}}^{\frac{1}{2}} = \frac{\rho}{\sqrt{1-\rho}} \left[\sum_{\ell=0}^{\infty} \left(\sum_{s,t \in \mathbb{Z}_+: s+t=\ell} \rho^{s+2t} \right) \tilde{\mathbf{A}}^\ell \mathbf{b} \right]_{\text{ew}}^{\frac{1}{2}} \\ &= \frac{\rho}{\sqrt{1-\rho}} \left[\sum_{\ell=0}^{\infty} \frac{\rho^{\ell+1} - \rho^{2\ell+2}}{\rho - \rho^2} \tilde{\mathbf{A}}^\ell \mathbf{b} \right]_{\text{ew}}^{\frac{1}{2}} \leq \frac{\rho}{\sqrt{1-\rho}} \left[\sum_{\ell=0}^{\infty} \frac{\rho^\ell}{1-\rho} \tilde{\mathbf{A}}^\ell \mathbf{b} \right]_{\text{ew}}^{\frac{1}{2}} = \frac{\rho}{1-\rho} \left[\left(\mathbf{I} + \sum_{\ell=1}^{\infty} \rho^\ell \tilde{\mathbf{A}}^\ell \right) \mathbf{b} \right]_{\text{ew}}^{\frac{1}{2}}. \end{aligned}$$

where the fourth inequality follows from Jensen's inequality provided $(1-\rho) \left(\mathbf{I} - \rho \tilde{\mathbf{A}} \right)^{-1}$ is a row-stochastic matrix. This concludes the proof.

We also remark that the bounds above also apply for each $t \geq 0$. Indeed, $\underline{\mu}(t) \leq \mathbf{q}(t)$, $\mu(t) \leq \bar{\mu}(t)$. Thus, it suffices to bound $\Delta\mu(t) = \bar{\mu}(t) - \underline{\mu}(t)$. An argument similar to (EC.6) shows that $|\Delta\mu(t)|_{\text{ew}} \leq \rho \tilde{\mathbf{A}} |\Delta\mu(t-1)|_{\text{ew}} + 2\delta$. Clearly, $|\Delta\mu(0)|_{\text{ew}} = 0 \leq 2 \left(\mathbf{I} - \rho \tilde{\mathbf{A}} \right)^{-1} \delta$. Assume that $|\Delta\mu(t-1)|_{\text{ew}} \leq 2 \left(\mathbf{I} - \rho \tilde{\mathbf{A}} \right)^{-1} \delta$. Then,

$$\begin{aligned} |\Delta\mu(t)|_{\text{ew}} &\leq \rho \tilde{\mathbf{A}} |\Delta\mu(t-1)|_{\text{ew}} + 2\delta \leq \rho \tilde{\mathbf{A}} \cdot 2 \left(\mathbf{I} - \rho \tilde{\mathbf{A}} \right)^{-1} \delta + 2\delta \\ &= - \left(\mathbf{I} - \rho \tilde{\mathbf{A}} \right) \cdot 2 \left(\mathbf{I} - \rho \tilde{\mathbf{A}} \right)^{-1} \delta + 2\delta + 2 \left(\mathbf{I} - \rho \tilde{\mathbf{A}} \right)^{-1} \delta = 2 \left(\mathbf{I} - \rho \tilde{\mathbf{A}} \right)^{-1} \delta, \end{aligned}$$

as desired. We conclude the proof. \square

In the following, we prove the corollary for Theorem 1.

Proof of Corollary 2: From Theorem 1, it holds that

$$\begin{aligned} \frac{1}{n} \|\mathbf{q}^* - \boldsymbol{\mu}^*\|_1 &\leq \frac{C_\rho}{n} \mathbf{e}^\top [\mathcal{C}(G; \rho)]_{\text{ew}}^{\frac{1}{2}} \\ &\stackrel{(a)}{\leq} \frac{C_\rho}{\sqrt{n}} \sqrt{\left\| [\mathcal{C}(G; \rho)]_{\text{ew}}^{\frac{1}{2}} \right\|_1} = \frac{C_\rho \sqrt{1-\rho}}{\sqrt{n}} \sqrt{\mathbf{e}^\top \left(\mathbf{I} + \sum_{\ell=1}^{\infty} \rho^\ell \tilde{\mathbf{A}}^\ell \right) \mathbf{b}}, \end{aligned}$$

where (a) follows due to Cauchy-Schwarz inequality. Further, (a) proves (9) in the corollary.

In the following, we will bound $\mathbf{e}^\top \left(\mathbf{I} + \sum_{\ell=1}^{\infty} \rho^\ell \tilde{\mathbf{A}}^\ell \right) \mathbf{b}$. Let us define $\mathbf{D} = \text{diag}(\mathbf{b})$, so that $\tilde{\mathbf{A}} = \mathbf{D}\mathbf{A}^\top$ holds, where \mathbf{A} is the adjacency matrix. Further, we define

$$\mathbf{Q}(s) := \mathbf{A}^\top \tilde{\mathbf{A}}^{s-1} = \mathbf{A}^\top (\mathbf{D}\mathbf{A}^\top)^{s-1}, \quad \forall s \geq 0.$$

Then, it holds that

$$\begin{aligned} \|\tilde{\mathbf{A}}^s \mathbf{b}\|_1 &= \mathbf{e}^\top \mathbf{D} \mathbf{Q}(s) \mathbf{D} \mathbf{e} \\ &= \sum_{i=1}^n \sum_{j=1}^n \frac{1}{d_i d_j} Q_{ij}(s) \leq \sum_{i=1}^n \sum_{j=1}^n \frac{1}{2} \left(\frac{1}{d_i^2} + \frac{1}{d_j^2} \right) Q_{ij}(s) = \frac{1}{2} \|\mathbf{D}^2 \mathbf{Q}(s) \mathbf{e}\|_1 + \frac{1}{2} \|\mathbf{Q}(s) \mathbf{D}^2 \mathbf{e}\|_1, \end{aligned} \quad (\text{EC.9})$$

where the inequality follows from the AM-GM inequality.

We then bound the two terms in (EC.9) as follows:

$$\frac{1}{2} \|\mathbf{D}^2 \mathbf{Q}(s) \mathbf{e}\|_1 = \frac{1}{2} \|\mathbf{D}^2 \mathbf{A}^\top \tilde{\mathbf{A}}^{s-1} \mathbf{e}\|_1 = \frac{1}{2} \|\mathbf{D} \mathbf{D} \mathbf{A}^\top \mathbf{e}\|_1 = \frac{1}{2} \|\mathbf{D} \mathbf{e}\|_1 = \frac{1}{2} \|\mathbf{b}\|_1,$$

where the second and the third equalities follow because $\tilde{\mathbf{A}}$ is row-stochastic, and

$$\frac{1}{2} \|\mathbf{Q}(s) \mathbf{D}^2 \mathbf{e}\|_1 = \frac{1}{2} \|(\mathbf{A}^\top \mathbf{D})^s \mathbf{D} \mathbf{e}\|_1 \leq \frac{1}{2} \|\mathbf{A}^\top \mathbf{D}\|_1^s \|\mathbf{D} \mathbf{e}\|_1 \stackrel{(b)}{=} \frac{1}{2} r^s(G) \|\mathbf{D} \mathbf{e}\|_1 = \frac{1}{2} r^s(G) \|\mathbf{b}\|_1,$$

where the inequality follows from the definition of matrix- ℓ_1 -norm as an operator norm and (b) follows since by inspection $r(G) = \mathbf{A}^\top \mathbf{D}$. Combined with (EC.9), we get

$$\|\tilde{\mathbf{A}}^s \mathbf{b}\|_1 \leq \frac{1}{2} (1 + r^s(G)) \|\mathbf{b}\|_1 \leq r^s(G) \|\mathbf{b}\|_1, \quad (\text{EC.10})$$

where the last inequality follows since by definition the largest out-in-degree ratio $r(G) \geq 1$. Therefore, as long as $\rho r(G) < 1$,

$$\mathbf{e}^\top \left(\mathbf{I} + \sum_{\ell=1}^{\infty} \rho^\ell \tilde{\mathbf{A}}^\ell \right) \mathbf{b} = \left\| \left(\mathbf{I} + \sum_{\ell=1}^{\infty} \rho^\ell \tilde{\mathbf{A}}^\ell \right) \mathbf{b} \right\|_1 \leq \|\mathbf{b}\|_1 + \sum_{\ell=1}^{\infty} \rho^\ell \|\tilde{\mathbf{A}}^\ell \mathbf{b}\|_1 \leq \frac{1}{1 - \rho r(G)} \|\mathbf{b}\|_1,$$

where the first inequality follows from the subadditivity of norms and the last inequality follows from (EC.10).

This concludes the proof. \square

Appendix C: Supporting Arguments for Section 4

For a refined upper bound on the approximation error, we first show Lemma EC.6, which is a refined version of Lemma 3.

LEMMA EC.6 (Improved Fixed-Point Deviation of Adoption Probability). *Under Assumptions 1, 2 and 3, for any diffusion instance $(G, \mathbf{v}, F_\epsilon(\cdot), \beta)$ and $t \geq 1$, we have*

$$\left| \mathbf{h}(\mathbf{q}(t-1)) - \mathbf{q}(t) \right|_{\text{ew}} \leq \frac{L_f \beta^2}{8} \mathcal{B}(G; \rho) \leq \frac{L_f \beta^2}{8} \left(\mathbf{I} - \rho^2 \tilde{\mathbf{A}} \right)^{-1} \mathbf{b}.$$

Proof of Lemma EC.6: Let $X_i = -v_i - \beta \frac{1}{d_i} \sum_{j \in \mathcal{N}_i} Y_j(t-1)$ and $\nu_i = -v_i - \beta \frac{1}{d_i} \sum_{j \in \mathcal{N}_i} q_j(t-1)$ for all $i \in V$. For any $i \in V$ and $t \geq 1$, the adoption probability of agent i at t can be written as

$$q_i(t) = \mathbb{E}_{\mathbf{Y}(t-1)} [\mathbb{E}[y_i(t) \mid \mathbf{Y}(t-1)]] = \mathbb{E}_{\mathbf{Y}(t-1)} \left[1 - F_\epsilon \left(-v_i - \beta \frac{\sum_{j \in \mathcal{N}_i} Y_j(t-1)}{d_i} \right) \right] = 1 - \mathbb{E}_{X_i} [F_\epsilon(X_i)].$$

With Assumption 3, we can apply Taylor expansion to $F_\epsilon(X_i)$ and get

$$\begin{aligned} \left| \mathbb{E}_{X_i} [F_\epsilon(X_i) - F_\epsilon(\nu_i)] \right| &= \left| \mathbb{E}_{X_i} \left[F_\epsilon(\nu_i) + f_\epsilon(\nu_i)(X_i - \nu_i) + \frac{1}{2} f'_\epsilon(\tilde{X}_i)(X_i - \nu_i)^2 - F_\epsilon(\nu_i) \right] \right| \\ &= \frac{1}{2} \left| \mathbb{E}_{X_i} \left[f'_\epsilon(\tilde{X}_i)(X_i - \nu_i)^2 \right] \right|, \end{aligned} \quad (\text{EC.11})$$

where \tilde{X}_i is a random variable such that \tilde{X}_i lies in between the random variable X_i and ν_i .

Consequently, we can upper bound (EC.11) by

$$\begin{aligned} \left| \mathbb{E}_{X_i} [F_\epsilon(X_i) - F_\epsilon(\nu_i)] \right| &= \frac{1}{2} \left| \mathbb{E}_{X_i} \left[f'_\epsilon(\tilde{X}_i)(X_i - \nu_i)^2 \right] \right| \leq \frac{1}{2} \mathbb{E}_{X_i} [|f'_\epsilon(\tilde{X}_i)|(X_i - \nu_i)^2] \leq \frac{L_f}{2} \text{Var}(X_i) \\ &= \frac{L_f \beta^2}{2} \text{Var} \left(\frac{1}{d_i} \sum_{j \in \mathcal{N}_i} Y_j(t-1) \right), \end{aligned}$$

where the first inequality comes from Jensen's inequality and the second inequality is from Assumption 3.

Let $\eta = \frac{L_f \beta^2}{8} \mathcal{B}(G; \rho)$. By applying Lemma 2, we can finally get

$$\left| \mathbb{E}[F_\epsilon(X_i) - F_\epsilon(\nu_i)] \right| \leq \eta,$$

which further leads to

$$1 - F_\epsilon \left(-v_i - \beta \frac{\sum_{j \in \mathcal{N}_i} q_j(t-1)}{d_i} \right) - \eta \leq q_i(t) \leq 1 - F_\epsilon \left(-v_i - \beta \frac{\sum_{j \in \mathcal{N}_i} q_j(t-1)}{d_i} \right) + \eta.$$

In conclusion, we have

$$\mathbf{h}(\mathbf{q}(t-1)) - \eta \leq \mathbf{q}(t) \leq \mathbf{h}(\mathbf{q}(t-1)) + \eta.$$

Moreover, because $\mathcal{B}(G; \rho) \leq (\mathbf{I} - \rho^2 \tilde{\mathbf{A}})^{-1} \mathbf{b}$ as proved in Lemma 2, we can also naturally have

$$\left| \mathbf{h}(\mathbf{q}(t-1)) - \mathbf{q}(t) \right|_{\text{ew}} \leq \frac{L_f \beta^2}{8} \mathcal{B}(G; \rho) \leq \frac{L_f \beta^2}{8} (\mathbf{I} - \rho^2 \tilde{\mathbf{A}})^{-1} \mathbf{b}.$$

□

Based on Lemma EC.6, we then show the proof for the refined Theorem 2 and Corollary 3.

Proof of Theorem 2: Following the same steps leading to (EC.8), with Lemma EC.6, we obtain

$$|\Delta \boldsymbol{\mu}|_{\text{ew}} \leq 2 (\mathbf{I} - \rho \tilde{\mathbf{A}})^{-1} \eta. \quad (\text{EC.12})$$

Therefore, following the same line of analysis in the proof of Theorem 1, it holds that

$$\begin{aligned} |\mathbf{q}^* - \boldsymbol{\mu}^*|_{\text{ew}} &\leq |\Delta \boldsymbol{\mu}|_{\text{ew}} \leq 2 (\mathbf{I} - \rho \tilde{\mathbf{A}})^{-1} \eta = \frac{L_f \beta^2}{4} (\mathbf{I} - \rho \tilde{\mathbf{A}})^{-1} \cdot \mathcal{B}(G; \rho) \\ &\leq \frac{L_f \beta^2}{4} (\mathbf{I} - \rho \tilde{\mathbf{A}})^{-1} \cdot (\mathbf{I} - \rho^2 \tilde{\mathbf{A}})^{-1} \mathbf{b} = \frac{L_f \beta^2}{4} \left[\left(\sum_{s=0}^{\infty} \rho^s \tilde{\mathbf{A}}^s \right) \left(\sum_{t=0}^{\infty} \rho^{2t} \tilde{\mathbf{A}}^t \right) \mathbf{b} \right] \\ &\leq \frac{L_f \beta^2}{4(1-\rho)} \cdot \left(\mathbf{I} + \sum_{\ell=1}^{\infty} \rho^\ell \tilde{\mathbf{A}}^\ell \right) \mathbf{b} = \frac{L_f \beta^2}{4(1-\rho)^2} \cdot \mathcal{C}(G; \rho). \end{aligned}$$

We conclude the proof. □

Proof of Corollary 3: By Theorem 2, we can upper bound the scaled ℓ_1 -norm as

$$\frac{1}{n} \|\mathbf{q}^* - \boldsymbol{\mu}^*\|_1 \leq \frac{\tilde{C}}{n} \mathbf{e}^\top \mathcal{C}(G; \rho) = \frac{(1-\rho)\tilde{C}}{n} \cdot \mathbf{e}^\top \left(\mathbf{I} + \sum_{\ell=1}^{\infty} \rho^\ell \tilde{\mathbf{A}}^\ell \right) \mathbf{b}$$

Following the proof of Corollary 2, the last term can be bounded by

$$\frac{(1-\rho)\tilde{C}}{n} \cdot \mathbf{e}^\top \left(\mathbf{I} + \sum_{\ell=1}^{\infty} \rho^\ell \tilde{\mathbf{A}}^\ell \right) \mathbf{b} \leq \frac{(1-\rho)\tilde{C}}{n(1-\rho r(G))} \|\mathbf{b}\|_1,$$

and we conclude the proof. \square

LEMMA EC.7. *Given an entrywise increasing sequence of vectors $\{\boldsymbol{\eta}(1), \boldsymbol{\eta}(2), \dots\}$, with $\lim_{t \rightarrow \infty} \boldsymbol{\eta}(t) = \boldsymbol{\eta}$, the dynamical system*

$$\boldsymbol{\mu}(t) = \mathbf{h}_{\boldsymbol{\eta}(t-1)}(\boldsymbol{\mu}(t-1)), \quad \forall t \geq 1,$$

converges to $\boldsymbol{\nu}^$ when t tends to infinity, where $\boldsymbol{\nu}^*$ is the fixed-point solution of the system of equations $\mathbf{h}_{\boldsymbol{\eta}}(\boldsymbol{\nu}) = \boldsymbol{\nu}$.*

Proof of Lemma EC.7: We first define a set of auxiliary dynamical systems, including an upper bound dynamical system $\boldsymbol{\nu}(t)$, and for any $\tau \geq 0$, a set of lower bound dynamical system $\boldsymbol{\mu}^{[\tau]}(t)$ as follows:

$$\begin{aligned} \nu_i(t) &= \begin{cases} q_i(0) & t=0 \\ 1 - F_\epsilon \left(-v_i - \beta \frac{\sum_{j \in \mathcal{N}_i} \nu_j(t-1)}{d_i} \right) + \eta_i & t>0 \end{cases}, \quad \text{for all } i \in V, \\ \mu_i^{[\tau]}(t) &= \begin{cases} q_i(0) & t=0 \\ 1 - F_\epsilon \left(-v_i - \beta \frac{\sum_{j \in \mathcal{N}_i} \mu_j^{[\tau]}(t-1)}{d_i} \right) + \eta_i(\min\{t-1, \tau\}) & t>0 \end{cases}, \quad \text{for all } i \in V. \end{aligned}$$

It is easy to verify that $\boldsymbol{\mu}^{[\tau]}(t) \leq \boldsymbol{\mu}(t) \leq \boldsymbol{\nu}(t)$. Then, by the contraction mapping theorem, we know that $\boldsymbol{\mu}^{[\tau]}(t)$ (resp. $\boldsymbol{\nu}(t)$) converges to $\boldsymbol{\mu}^{[\tau]*}$ (resp. $\boldsymbol{\nu}^*$) where $\boldsymbol{\mu}^{[\tau]*}$ (resp. $\boldsymbol{\nu}^*$) is the fixed-point solution for the system of equations $\mathbf{h}_{\boldsymbol{\eta}(\tau)}(\boldsymbol{\mu}^{[\tau]}) = \boldsymbol{\mu}^{[\tau]}$ (resp. $\mathbf{h}_{\boldsymbol{\eta}}(\boldsymbol{\nu}) = \boldsymbol{\nu}$). Thus, the following result holds,

$$\boldsymbol{\mu}^{[\tau]*} \leq \boldsymbol{\mu}^* \leq \boldsymbol{\nu}^*.$$

By the proof of Theorem 1, we know that the sequence $\{\boldsymbol{\mu}^{[0]*}, \boldsymbol{\mu}^{[1]*}, \dots\}$ is also entrywise increasing. Hence, by monotone convergence theorem, we have

$$\lim_{\tau \rightarrow \infty} \boldsymbol{\mu}^{[\tau]*} = \boldsymbol{\nu}^*.$$

Finally, by squeeze theorem, we can conclude that $\boldsymbol{\mu}^* = \boldsymbol{\nu}^*$. \square

Proof of Theorem 3: We use the diffusion instance given in the main text to show the lower bound. We first remark on the following facts that will be used in the next. For this specific instance, the CDF F_ϵ , PDF f_ϵ and the derivative of PDF f'_ϵ are given by

$$F_\epsilon(x) = \frac{1}{1 + e^{-x}}, \quad f_\epsilon(x) = \frac{e^{-x}}{(1 + e^{-x})^2}, \quad \text{and} \quad f'_\epsilon(x) = \frac{e^{-x}(e^{-x} - 1)}{(1 + e^{-x})^3}.$$

To lower bound the variance, we first provide a lower bound of the variance of the adoption state $Y_i(t)$ for each agent $i \in V$ and $t \geq 1$. Since $\text{Var}(Y_i(t)) = q_i(t)(1 - q_i(t))$ where $q_i(t) = \mathbb{E}[1 - F_\epsilon(-v - \beta \frac{1}{d_i} \sum_{j \in \mathcal{N}_i} Y_j(t-1))]$, we can derive that, when the intrinsic value of all agents $v_i = v = -\beta - \xi$,

$$\text{Var}(Y_i(t)) \geq [1 - F_\epsilon(-v)] \cdot F_\epsilon(-v - \beta) = [1 - F_\epsilon(\xi + \beta)] \cdot F_\epsilon(\xi). \quad (\text{EC.13})$$

Then, we will establish the lower bound on the covariance matrix of adoption states, which is similar to the proof of Lemma 1. Specifically, we show that $\text{Cov}_{\mathbf{Y}(t-1)} \left(\mathbb{E}_{\epsilon(t)} [Y_i(t) \mid \mathbf{Y}(t-1)], \mathbb{E}_{\epsilon(t)} [Y_{i'}(t) \mid \mathbf{Y}(t-1)] \right)$ can be lower bounded recursively as

$$\begin{aligned} & \text{Cov}_{\mathbf{Y}(t-1)} \left(\mathbb{E}_{\epsilon(t)} [Y_i(t) \mid \mathbf{Y}(t-1)], \mathbb{E}_{\epsilon(t)} [Y_{i'}(t) \mid \mathbf{Y}(t-1)] \right) \\ &= \text{Cov} \left(1 - F_\epsilon \left(-v_i - \beta \frac{\sum_{j \in \mathcal{N}_i} Y_j(t-1)}{d_i} \right), 1 - F_\epsilon \left(-v_{i'} - \beta \frac{\sum_{j' \in \mathcal{N}_{i'}} Y_{j'}(t-1)}{d_{i'}} \right) \right) \\ &\geq \frac{(\ell\beta)^2}{d_i d_{i'}} \text{Cov} \left(\sum_{j \in \mathcal{N}_i} Y_j(t-1), \sum_{j' \in \mathcal{N}_{i'}} Y_{j'}(t-1) \right) \end{aligned} \quad (\text{EC.14a})$$

$$= \frac{\rho_\ell^2}{d_i d_{i'}} \sum_{j \in \mathcal{N}_i} \sum_{j' \in \mathcal{N}_{i'}} \text{Cov}(Y_j(t-1), Y_{j'}(t-1)). \quad (\text{EC.14b})$$

where (EC.14a) follows from Corollary EC.1 and (EC.14b) follows from the definition of covariance of linear combinations and the fact that $\rho_\ell = \ell\beta$.

When $t = 1$, the covariance matrix can be lower bounded by $c\mathbf{I}$. Therefore, we can lower bound the covariance matrix $\Sigma(t)$ iteratively as

$$\Sigma(t) \geq c\mathbf{I} + \rho_\ell^2 \tilde{\mathbf{A}} \Sigma(t-1) \tilde{\mathbf{A}}^\top \geq c\mathbf{I} + \rho_\ell^2 \tilde{\mathbf{A}} \left[c\mathbf{I} + \rho_\ell^2 \tilde{\mathbf{A}} \Sigma(t-2) \tilde{\mathbf{A}}^\top \right] \tilde{\mathbf{A}}^\top \geq \dots \geq c \left[\mathbf{I} + \sum_{\tau=1}^{t-1} \rho_\ell^{2\tau} \tilde{\mathbf{A}}^\top \left(\tilde{\mathbf{A}}^\top \right)^\tau \right],$$

where the inequalities follow since (EC.13) and (EC.14).

Furthermore, we have

$$\kappa(t) \geq c \left[\sum_{\tau=1}^t \rho_\ell^{2\tau-2} \tilde{\mathbf{A}}^\top \odot \left(\tilde{\mathbf{A}}^\top \right)^\tau \right] \mathbf{e}.$$

We then bound the difference between $\mathbf{q}(t)$ and $\mathbf{h}(\mathbf{q}(t-1))$. Similar to proof of Lemma EC.6, we obtain

$$\begin{aligned} \mathbb{E}[F_\epsilon(X_i) - F_\epsilon(\nu_i)] &= \frac{1}{2} \mathbb{E}[f'_\epsilon(\tilde{X}_i)(X_i - \nu_i)^2] \leq -\frac{\ell_f}{2} \text{Var}(X_i) = -\frac{\ell_f \beta^2}{2} \text{Var} \left(\frac{1}{d} \sum_{j \in \mathcal{N}_i} Y_j(t-1) \right) \\ &\leq -\frac{\ell_f \beta^2 c}{2} \left[\sum_{\tau=1}^{t-1} \rho_\ell^{2\tau-2} \tilde{\mathbf{A}}^\top \odot \left(\tilde{\mathbf{A}}^\top \right)^\tau \right] \mathbf{e}, \end{aligned}$$

where the first inequality follows because $\tilde{X}_i \in [\xi, \xi + \beta]$ and $f'_\epsilon(x) < -\ell_f$ for x in this range. Therefore,

$$\begin{aligned} q_i(t) &= 1 - \mathbb{E}[F_\epsilon(X_i)] \geq 1 - F_\epsilon(\nu_i) + \frac{\ell_f \beta^2 c}{2} \left[\sum_{\tau=1}^t \rho_\ell^{2\tau-2} \tilde{\mathbf{A}}^\top \odot \left(\tilde{\mathbf{A}}^\top \right)^\tau \right] \mathbf{e} \\ &= h(\mathbf{q}(t-1))_i + \frac{\ell_f \beta^2 c}{2} \left[\sum_{\tau=1}^t \rho_\ell^{2\tau-2} \tilde{\mathbf{A}}^\top \odot \left(\tilde{\mathbf{A}}^\top \right)^\tau \right] \mathbf{e}. \end{aligned}$$

In the vector form, we have

$$\mathbf{q}(t) \geq \mathbf{h}(\mathbf{q}(t-1)) + \frac{\ell_f \beta^2 c}{2} \left[\sum_{\tau=1}^t \rho_\ell^{2\tau-2} \tilde{\mathbf{A}}^\top \odot \left(\tilde{\mathbf{A}}^\top \right)^\tau \right] \mathbf{e}. \quad (\text{EC.15})$$

Finally, we lower bound the approximation error in a way analogous to Theorem 1. Let $\eta(t) = \frac{\ell_f \beta^2 c}{2} \left[\sum_{\tau=1}^t \rho_\ell^{2\tau-2} \tilde{\mathbf{A}}^\top \odot \left(\tilde{\mathbf{A}}^\top \right)^\tau \right] \mathbf{e}$. Let

$$\bar{\mu}_i(t) = \begin{cases} q_i(0) & t = 0 \\ 1 - F_\epsilon \left(-v_i - \beta \frac{\sum_{j \in \mathcal{N}_i} \bar{\mu}_i(t-1)}{d_i} \right) + \eta(t-1)_i & t > 0 \end{cases}, \quad \text{for all } i \in V.$$

We will show $\mathbf{q}(t) \geq \bar{\boldsymbol{\mu}}(t) \geq \boldsymbol{\mu}(t)$ by induction.

Base case $t = 0$: By definition, we have $\mathbf{q}(0) = \bar{\boldsymbol{\mu}}(0) = \boldsymbol{\mu}(0)$.

To Show $t = s + 1$: Assume that $\mathbf{q}(s) \geq \bar{\boldsymbol{\mu}}(s) \geq \boldsymbol{\mu}(s)$. We have

$$\mathbf{q}(s+1) \geq \mathbf{h}_{\delta(s)}(\mathbf{q}(s)) \geq \mathbf{h}_{\delta(s)}(\bar{\boldsymbol{\mu}}(s)) = \bar{\boldsymbol{\mu}}(s+1) \geq \mathbf{h}_{\delta(s)}(\boldsymbol{\mu}(s)) \geq \mathbf{h}(\boldsymbol{\mu}(s)) = \boldsymbol{\mu}(s+1),$$

where the first inequality follows from (EC.15), the second and third inequalities follow Proposition 2 and the induction hypothesis, and the last inequality is trivial because $\boldsymbol{\eta}(s) > 0$.

While the dynamical system $\{\bar{\boldsymbol{\mu}}(t)\}$ is not a fixed-point iteration, we can show that it still converges. Let $\boldsymbol{\eta} = \lim_{t \rightarrow \infty} \boldsymbol{\eta}(t) = \frac{\ell_f \beta^2 c}{2} \mathcal{B}(G; \rho_\ell)$. By Lemma EC.7, we can show that $\{\bar{\boldsymbol{\mu}}(t)\}$ converges to $\boldsymbol{\nu}^*$, where $\boldsymbol{\nu}^*$ is the fixed-point solution of $\mathbf{h}_\eta(\boldsymbol{\nu}) = \boldsymbol{\nu}$.

Thus, the following result holds,

$$\mathbf{q}^* \geq \bar{\boldsymbol{\mu}}^* = \boldsymbol{\nu}^* \geq \boldsymbol{\mu}^*, \quad (\text{EC.16})$$

where $\bar{\boldsymbol{\mu}}^*$ is the limit of $\bar{\boldsymbol{\mu}}(t)$. Therefore, we know that $\mathbf{q}^* - \boldsymbol{\mu}^*$ will be lower bounded by $\bar{\boldsymbol{\mu}}^* - \boldsymbol{\mu}^*$. We have

$$\bar{\boldsymbol{\mu}}^* - \boldsymbol{\mu}^* = \mathbf{h}(\bar{\boldsymbol{\mu}}^*) - \mathbf{h}(\boldsymbol{\mu}^*) + \boldsymbol{\eta} \geq \rho_\ell \tilde{\mathbf{A}} (\bar{\boldsymbol{\mu}}^* - \boldsymbol{\mu}^*) + \boldsymbol{\eta},$$

where the inequality follow from Assumption 3.

As a result, we can arrive at the lower bound

$$\bar{\boldsymbol{\mu}}^* - \boldsymbol{\mu}^* \geq \left(\mathbf{I} - \rho_\ell \tilde{\mathbf{A}} \right)^{-1} \boldsymbol{\eta} = \frac{\ell_f \beta^2 c}{2} \left(\mathbf{I} - \rho_\ell \tilde{\mathbf{A}} \right)^{-1} \mathcal{B}(G; \rho_\ell),$$

and conclude the proof. \square

Appendix D: Supplementary Numerical Experiments on the FPA Scheme

D.1. Numerical Experiments on the 10-Node Example Instance

We construct an undirected network comprising 10 nodes. The network structure is visualized in Fig. 1a, while the intrinsic values assigned to each agent are detailed in Table EC.1. We set the network effect intensity at $\beta = 3.5$ and assume that the random noise distribution is $\epsilon_i(t) \stackrel{\text{i.i.d.}}{\sim} \text{Logistic}(0, 1)$ for all $i \in V$ and $t \geq 0$. The characteristics of this example network, along with the numerical results obtained from different models, are presented in Table EC.1.

D.1.1. Discussions on Assumption 2. Although assumptions similar to Assumption 2 are common in the network literature, its implications for the FPA scheme deserve further exploration. The parameter ρ has two elements, namely the network effect intensity β and the Lipschitz constant L of $F_\epsilon(\cdot)$. We conduct experiments by varying each of these two components, benchmarking against the misspecified model where the network effect is not incorporated (See MM in Section 3.1). Figure EC.1 shows the MAPE of the FPA scheme with different values of β and L .

(i) Sensitivity analysis of network effect intensity. We vary β from 0 to 10 and keep other parameters fixed. Given $\epsilon_i \stackrel{\text{i.i.d.}}{\sim} \text{Logistic}(0, 1)$ with $L = 1/4$, our experiments cover both cases when Assumption 2 is satisfied

Table EC.1 Characteristics and results of the 10-node example instance

Node	In-degree d	Intrinsic value v	q^*	μ^*	FPA error	q^{MM}	MM error
0	5	-1.7064	0.5126	0.5292	0.0166	0.1536	-0.3590
1	7	-1.2453	0.5932	0.6069	0.0137	0.2235	-0.3697
2	4	-0.8789	0.6325	0.6524	0.0199	0.2934	-0.3391
3	4	-3.9454	0.1442	0.1221	-0.0222	0.0190	-0.1253
4	3	-0.0822	0.7827	0.8219	0.0393	0.4795	-0.3032
5	5	-3.4441	0.1933	0.1731	-0.0202	0.0309	-0.1624
6	3	-0.2877	0.7341	0.7755	0.0414	0.4286	-0.3055
7	2	-2.9084	0.3287	0.2849	-0.0438	0.0517	-0.2770
8	2	-1.2859	0.6702	0.7646	0.0944	0.2166	-0.4536
9	1	-0.6963	0.7416	0.8786	0.1371	0.3326	-0.4090

Note: p^* is calculated by first constructing a 1,024-state MC according to Section 2.2 and calculating the stationary distribution. μ^* is calculated by conducting fixed-point iteration according to (5), and FPA error equals $(\mu_i^* - q_i^*)$. p^{MM} is calculated as $\mathbb{E}[\mathbb{1}\{v_i + \epsilon_i \geq 0\}]$, and MM error equals $(\mu_i^{\text{MM}} - q_i^*)$.

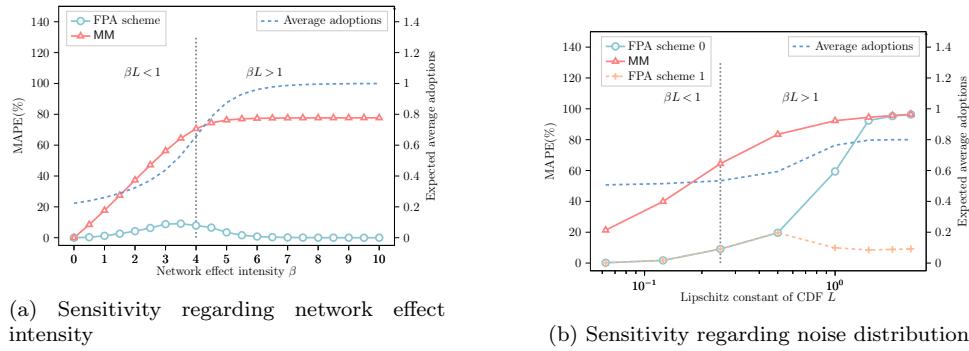


Figure EC.1 Sensitivity of the approximation error against parameter ρ . In the right subfigure, the horizontal axis is in log scale. The FPA scheme 0 (1, resp.) represents the FPA solution initialized with $\mu(0) = 0$ ($\mu(0) = 1$, resp.).

and violated. As shown in Figure EC.1a, the MAPE first increases with β , but at a notably slower rate compared to MM. When $\rho = \beta L > 1$, the MAPE gradually declines to 0.

For $\rho < 1$, it is not surprising that the FPA scheme performs exceptionally well when β is close to 0 (i.e., the network effect is weak). As β increases, the network effect becomes more influential, slightly degrading FPA performance. This dependence is reflected in constants C_ρ and \tilde{C} in our theoretical results. Nevertheless, even at $\beta = 3.5$, where the MAPE peaks at 9.11%, FPA performance remains commendable, substantially lower than that of MM (64.42%). This underscores the robustness of the FPA scheme, even in the presence of strong network effects in a small network.

When $\rho > 1$, we observe an intriguing trend: the MAPE decreases as β increases. This is because L is not uniformly tight for $F_\epsilon(\cdot)$. With large β , the network effects heavily influence user behavior, pushing adoption probabilities close to 1 for many agents. Thus, the nominal utility in the FPA scheme, given by $v_i + \beta \frac{\sum_{j \in \mathcal{N}_i} \mu_j(t-1)}{d_i}$, gravitates towards flat areas of the CDF, where the effective Lipschitz constant is much smaller than L . Even if the nominal utility occasionally falls into regions with larger Lipschitz constants, thereby violating Assumption 2, the FPA scheme remains resilient.

(ii) Sensitivity analysis of noise distribution. The experiment above may lead readers to conceive that Assumption 2 is conservative. However, we demonstrate that this is not the case by varying L . Particularly, we assess FPA performance for Logistic distribution with different parameters, where $\epsilon_i \stackrel{\text{i.i.d.}}{\sim} \text{Logistic}(0, s)$ and s ranges from 0.0625 to 2.5. The associated Lipschitz constant is given by $L = \frac{1}{4s}$, and we hold other parameters fixed. In Figure EC.1b, we observe a continuous increase in MAPE as L increases. Notably, when $\rho > 1$, FPA performance deteriorates drastically, approaching that of MM. This deterioration results from the violation of Proposition 2, e.g., there may be multiple solutions to the FPA scheme. In particular, the FPA solution initialized at $\mu(0) = \mathbf{0}$ diverges from \mathbf{q}^* . To offer a more comprehensive view, we also present an additional FPA solution initialized at $\mu(0) = \mathbf{1}$ in Figure EC.1b. These two solutions exhibit divergent performance, with the latter significantly outperforming the former when $\rho > 1$. We remark that other FPA solutions could exist with different initial values, and it is challenging to determine which solution will perform best *a priori*.

Upon closer examination, a key difference emerges between these two scenarios. In the first, the intrinsic utility \mathbf{v} retains its relative position within the noise distribution. In contrast, in the second, a small s causes v_i to vary widely, pushing some agents strongly toward adoption and others toward non-adoption. When Assumption 2 is violated in this context, the FPA scheme may admit multiple solutions, complicating its use. In short, the impacts of violating Assumption 2 are multifaceted and context-dependent. We leave further explorations of these phenomena to future studies.

D.2. Numerical Experiments on Highly-Structured Symmetric Networks

To illustrate the exact performance of the FPA scheme, we focus on two kinds of highly-structured symmetric networks, namely directed star networks and complete networks. These simple and symmetric structures make it easier to calculate the limiting adoption probability. We further simplify the diffusion instance by setting the intrinsic value of all agents in the network to be the same as v . This allows us either to directly compute the limiting adoption probability or to construct an MC with a much smaller state space.

Network instances:

- *Directed star networks.* A star network consists of a central node and several surrounding nodes. We consider the directed version where the edges only point from surrounding nodes to the central node, not vice versa. Figure EC.2a shows an example of network size $n = 6$.
- *Complete networks.* A complete network is the network where each node is directly connected to every other node. Figure EC.2b shows an example of network size $n = 6$.

For directed star networks, the adoption decisions of surrounding nodes are independent of each other. Therefore, we can directly calculate the limiting adoption probability of the central node as

$$q = \sum_{i=0}^{n-1} \binom{n-1}{i} (1 - F_\epsilon(-v))^i F_\epsilon(-v)^{n-1-i} \cdot \left[1 - F_\epsilon\left(-v - \beta \frac{i}{n-1}\right) \right].$$

For complete networks, we can construct a more efficient MC by using the number of adopted agents as the MC states, rather than considering the combination of all agents' adoption states. The transition probability of this MC can be defined as

$$P(i, j) = \sum_{k=0}^{\min\{i,j\}} \binom{i}{k} \left[1 - F_\epsilon\left(-v - \beta \frac{i-1}{n-1}\right) \right]^k F_\epsilon\left(-v - \beta \frac{i-1}{n-1}\right)^{i-k} \binom{n-i}{j-k}$$

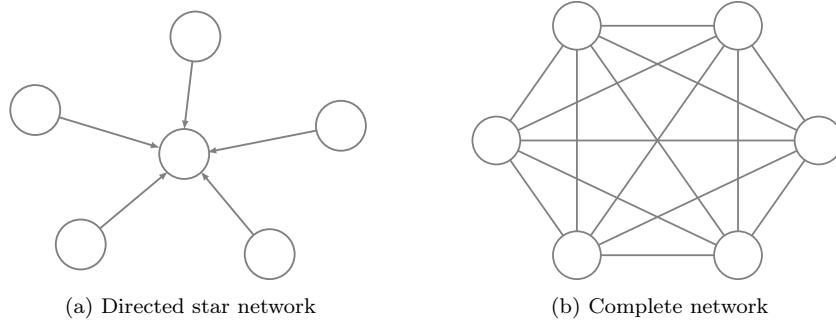


Figure EC.2 Illustration of highly structured symmetric network structure.

$$\cdot \left[1 - F_\epsilon \left(-v - \beta \frac{i}{n-1} \right) \right]^{j-k} F_\epsilon \left(-v - \beta \frac{i}{n-1} \right)^{n-i-j+k}.$$

Hence, the limiting adoption probability for this MC can be directly calculated.

We measure FPA performance by the percentage error (PE) of the representative node, given in the following equation:

$$\text{PE} = \frac{\mu_i^* - q_i^*}{q_i^*} \cdot 100\%.$$

In directed star networks, we focus solely on the central node because the surrounding nodes have zero in-degree and can thus be perfectly approximated by the FPA scheme. In complete networks, the PE is identical for all nodes. Therefore, the PE for any arbitrary node in a complete network is equivalent to the mean average percentage error.

To assess FPA performance, we investigate two scenarios for both types of network structures: (i) a sequence of diffusion instances with different intrinsic values, and (ii) a sequence of diffusion instances with different network sizes. For these experiments, we set the network effect intensity to be $\beta = 1$ and generate the random noise $\epsilon_i(t) \stackrel{\text{i.i.d.}}{\sim} \text{Logistic}(0, 1)$.

(i) The accuracy with regard to intrinsic values. We choose the intrinsic value v from -5 to 5 in increments of 0.1. These instances are tested on networks of size $n \in \{10, 20, 30\}$. Figure EC.3 shows the PE of both network structures at different intrinsic values. Overall, all instances have a small absolute percentage error

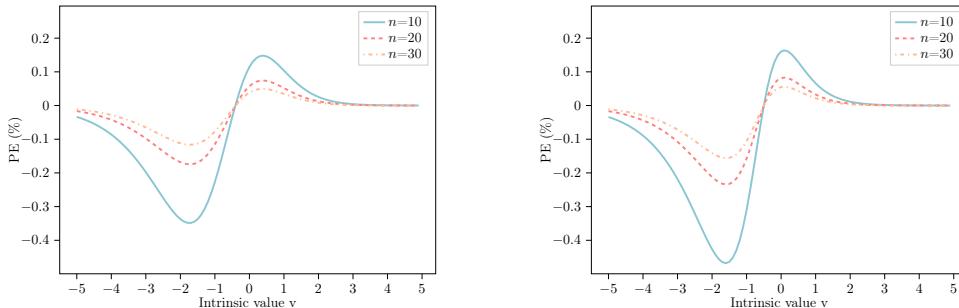


Figure EC.3 PE versus intrinsic value. Left: Directed star network; Right: Complete network.

(less than 0.5%), illustrating the high accuracy of the FPA solution. We notice that the PE curves of different network structures possess similar shapes, however, the exact values are slightly different. In general, the FPA

scheme tends to underestimate the adoption probability when the intrinsic values are small and overestimate it when they are large. There exist two critical points at around $v = -1.7$ and $v = 0.4$ where the PE reaches extremes. These points exhibit the worst cases and align with regions where the CDF $F_\epsilon(\cdot)$ has the largest curvature.

(ii) *The accuracy with regard to network size.* We then focus on instances with intrinsic values at the two previously mentioned critical points $v \in \{-1.7, 0.4\}$. We choose the network size n from 2 to 50. Figure EC.4 shows the PE across these different network sizes. Regardless of the network structure and the intrinsic

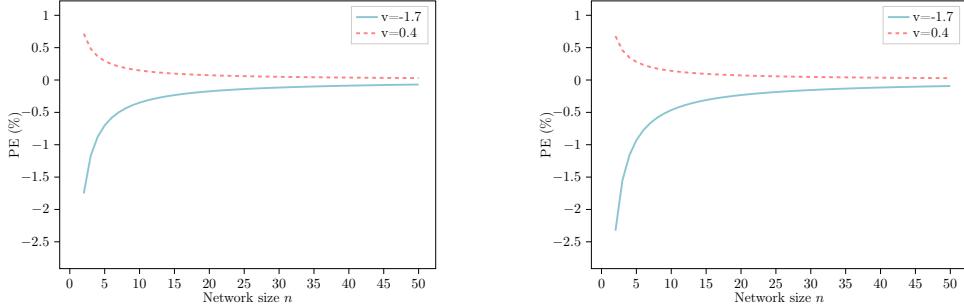


Figure EC.4 PE versus network size. Left: Directed star network; Right: Complete network.

values, PE converges to 0 rapidly when the network size increases. This can be theoretically confirmed, and we have explicitly demonstrated it in Corollary 1 or a refined version in Theorem 2. Our findings show that, for highly-structured networks, the FPA scheme offers excellent approximation quality and exhibits asymptotic convergence as the network size grows.

D.3. Agent-Based Simulation Settings

In this section, we discuss the simulation methods used in our work, which serve as benchmarks for comparing the FPA scheme. First, we describe the naïve ABS method, which is used both for benchmarking and establishing the underlying ground truth. We also empirically examine the warm-up period for the simulation. Then, we will introduce the more advanced A-ABS method.

D.3.1. Naïve ABS Method. The simulation starts with all agents in a non-adopted state. At each time step, the state of each agent is updated based on the previous time step using the agent behavior model. Specifically, for agent i at time t , we determine their adoption state based on $\mathbf{Y}(t-1)$ using the following equation:

$$Y_i(t) = \mathbb{1} \left\{ v_i + \beta \frac{\sum_{j \in \mathcal{N}_i} Y_j(t-1)}{d_i} + \epsilon_i(t) \geq 0 \right\}.$$

We designate the first 1,000 time steps as the warm-up period to allow the system to reach a steady state. These initial steps are discarded from the analysis to avoid transient bias. After the warm-up period, we collect T time steps of the adoption states. Following the first equation in (3), the adoption probability for agent i is estimated as:

$$\hat{q}_i = \frac{1}{T} \sum_{t=1}^T Y_i(t).$$

For the ground truth, we set the simulation run length to 100,000 steps beyond the warm-up period. When testing the performance of ABS methods, we stop the simulation once the real-time MAPE of ABS methods falls below that of the FPA scheme.

Additionally, we conduct experiments to empirically demonstrate why the warm-up period can be set as 1,000 time steps. Rather than focusing on individual states of the MC, we track the average cumulative adoption proportion across the entire population as an indicator, which is given by:

$$\frac{1}{T} \sum_{t=1}^T \frac{1}{n} \sum_{i \in V} Y_i(t).$$

In Figure EC.5, we show how the average cumulative adoption proportion evolves over time. We test on 4 different diffusion instances, each represented by a randomly sampled Erdős-Rényi network $G(n, p(n))$. We choose the network size from $n \in \{10, 100, 1000, 10000\}$ and keep the probability of edge existence to be $p(n) = 0.1$. We set the network effect intensity to be $\beta = 1$ and generate the random noise $\epsilon_i(t) \stackrel{\text{i.i.d.}}{\sim} \text{Logistic}(0, 1)$. Our results show that after 1,000 time steps, all tested trajectories reach a steady state. Larger networks

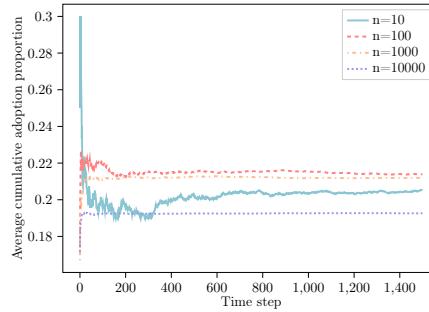


Figure EC.5 Average cumulative adoption proportion versus the time steps of Naïve ABS

tend to reach a steady state more quickly. Additional tests with varying parameters yielded similar results, leading us to conclude that a warm-up period of 1,000 time steps is sufficient for our context.

D.3.2. Accelerated ABS method. The Naïve ABS method is computationally expensive due to the large number of time steps required to move past the transient phase and the need to sample many adoption states for accurate estimation. Furthermore, because the adoption state $Y_i(t)$ is correlated over time and across the network, the time required to achieve a target accuracy becomes even more substantial.

We observe that the adoption state $Y_i(t)$ is a compound Bernoulli variable, with uncertainty in estimating the limiting adoption probability \mathbf{q}^* arising from two sources: (i) the neighbors' adoption states $\mathbf{Y}(t-1)$ and (ii) the random noise $\epsilon(t)$. The Naïve ABS method incorporates both sources of uncertainty, which require large samples for an accurate estimation.

Inspired by the second equation in (3), we leverage the analytical form of the adoption probability to improve sampling efficiency. In particular, we reduce one source of uncertainty. Given $\mathbf{Y}(t-1)$, we can express the adoption probability $q_i(t)$ as follows:

$$\begin{aligned} q_i(t) &= \mathbb{E}_{\mathbf{Y}(t-1)} \mathbb{E}_{\epsilon(t)} [Y_i(t) | \mathbf{Y}(t-1)] = \mathbb{E}_{\mathbf{Y}(t-1)} \left[\mathbb{E}_{\epsilon(t)} \left[\mathbb{1} \left\{ v_i + \beta \frac{\sum_{j \in \mathcal{N}_i} Y_j(t-1)}{d_i} + \epsilon_i(t) \geq 0 \right\} \middle| \mathbf{Y}(t-1) \right] \right] \\ &= \mathbb{E}_{\mathbf{Y}(t-1)} \left[1 - F_\epsilon \left(-v_i - \beta \frac{\sum_{j \in \mathcal{N}_i} Y_j(t-1)}{d_i} \right) \right]. \end{aligned}$$

Since $\mathbf{Y}(t-1)$ is observable during the simulation, we can estimate the adoption probability more efficiently as follows:

$$\hat{q}_i = \frac{1}{T} \sum_{t=1}^T \left[1 - F_\epsilon \left(-v_i - \beta \frac{\sum_{j \in \mathcal{N}_i} Y_j(t-1)}{n_i} \right) \right].$$

This A-ABS approach reduces the computational burden by replacing the need for large samples of adoption states with conditional probability evaluations, improving the efficiency of the simulation process.

D.4. Numerical Experiments for Power-Law Networks

We extend our examination of the FPA scheme to another important class of random networks, i.e., power-law networks. This class of networks exhibits a degree distribution that follows a power-law pattern. We consider a sequence of directed power-law networks with n nodes and define the associated CDF of the degree distribution as $F_d(\cdot; n)$. The network in-degrees and out-degrees are generated using the following CDF:

$$F_d(x; n) = \mathbb{P}(d \leq x) = \frac{1 - \left(\frac{x}{d_{\min}} \right)^{1-\alpha}}{1 - \left(\frac{d_{\max}}{d_{\min}} \right)^{1-\alpha}} \quad \text{for } d_{\min} \leq x \leq d_{\max} = n,$$

where α is the exponent of power-law distribution. We set d_{\min} to be 2 and d_{\max} to be n . Correspondingly, the probability mass function satisfies $f_d(x) \propto x^{-\alpha}$, which aligns with the conventional definition of a power-law distribution. Power-law networks often pose significant challenges for the analysis and optimization on networks due to the prevalence of low-degree nodes in such networks. Focusing on power-law networks with $d_{\min} = 2$ and $\rho = 0.875$ (see Section 5) allows us to test the limit of the FPA scheme.

We generate power-law network based on $F_d(\cdot; n)$ following the approach proposed by Huang et al. (2022). Detailed information on the generation process is included in Appendix D.5 for completeness. In this generation process we use an auxiliary parameter θ to account for the pairwise correlation between the in-degree and out-degree sequences. Although θ is not the exact correlation between these two sequences, it approximates the actual correlation between in-degrees and out-degrees, particularly for large values of n .

We conduct two sets of experiments to test the FPA scheme across different power-law exponents α and pairwise correlations θ . For each parameter combination, we conduct 100 repetitions to ensure stable performance metrics. The results are presented in Figure EC.6. In general, The FPA scheme still performs reasonably well. Also, we observe a consistent decrease in the MAPE as n increases across all tested parameter pairs, albeit at a relatively modest pace in comparison with Erdős-Rényi networks. Additionally, power-law networks exhibit both a higher mean and greater variance in MAPE. The increased mean MAPE in power-law networks is largely attributable to a higher proportion of low in-degree nodes. The increased variance, on the other hand, is primarily due to the more intricate structural variations inherent to power-law networks when specified parameters are used.

We note that the power-law exponent α has a crucial impact on the degree distribution. Typically, power-law networks feature an $\alpha > 2$ to avoid divergence in the expected degree. In the case where $\alpha = 3$, the network adheres to a model generated through the preferential attachment process. Accordingly, we select α from the set $\{2.5, 3, 3.5\}$ and set $\theta = 0$ to construct Figure EC.6a. It is evident that as α increases, the

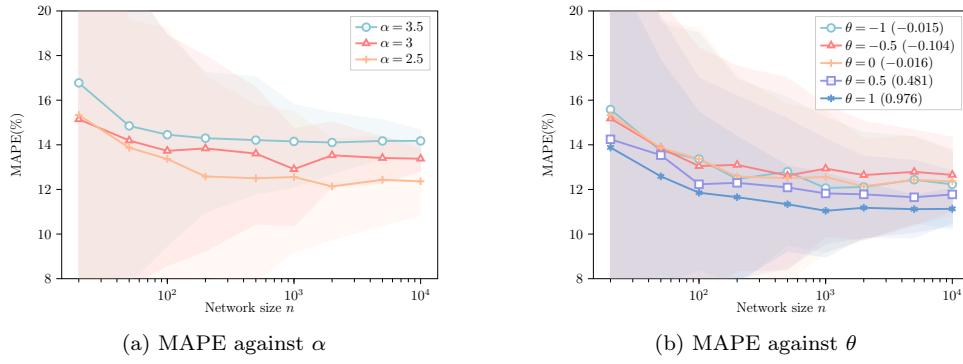


Figure EC.6 FPA Performance on power-law networks for different α and θ values. All horizontal axes are in the log scale. Shaded areas represent the 95% confidence interval. In the legend of the right subfigure, numbers within parentheses represent empirical correlations between in-degrees and out-degrees.

proportion of low-degree nodes also increases. Consistent with our theoretical analysis, we observe that the MAPE tends to rise as α increases.

In the second experiment, we generate the in-degree and out-degree sequences with $\theta \in \{-1, -0.5, 0, 0.5, 1\}$. Corollary 2 shows that the FPA scheme's performance is related to the imbalance level of the network, which can be captured by this pairwise correlation coefficient θ . Specifically, a large positive θ indicates a strong positive correlation between in-degree and out-degree sequences, resulting in a more balanced network. Conversely, a negative θ , suggests a more imbalanced network. From Figure EC.6b, the MAPE remains relatively stable when θ ranges between -1 and 0. However, it substantially diminishes as θ becomes positive, which aligns with our theoretical findings that the FPA scheme performs worse on more imbalanced networks.

Finally, we focus on FPA performance for nodes with different in-degrees d . For illustration, we choose instances with $\alpha = 2.5$ and $\theta = 0$. In Figure EC.7a, we illustrate how the MAPE varies with respect to the in-degree d . We can notice the MAPE consistently decreases for nodes with $d \geq 1$ as d increases. For nodes

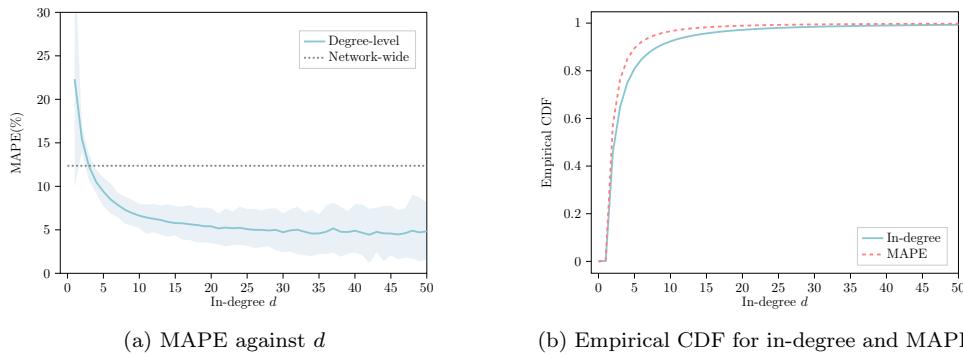


Figure EC.7 FPA Performance on power-law networks with regard to the in-degree values. The dotted line in the left subfigure is the average network-wide MAPE. Shaded area represents 95% confidence interval.

with more than 10 in-neighbors, the MAPE diminishes to less than 6.5%. Furthermore, the network-wide MAPE stands at approximately 12.36%; notably, only nodes with an in-degree of less than 3 exhibit errors above this level. In Figure EC.7b, we extend our analysis by displaying the empirical CDF for the in-degree

distribution and the MAPE. Given the nature of power-law networks, a substantial number of nodes exhibit low in-degree. Moreover, we can notice that the empirical CDF of the node-level MAPE is higher than that of in-degree, indicating these low in-degree nodes are also associated with larger errors. Specifically, 84.89% of the total error is attributable to agents with fewer than 5 in-neighbors, and 95.97% of the error can be attributed to agents with fewer than 10 in-neighbors.

D.5. Supplementary Discussions on Random Networks.

In our numerical experiments of random networks, we generate our data following the setup outlined in Huang et al. (2022), which also offers an excellent discussion on the key properties of this type of network. In the following, we revisit some of the discussions on parameter selection and instance construction for both Erdős-Rényi and power-law networks, supplementing them with additional numerical illustrations for more robust empirical support. For more details, please refer directly to this paper.

(i) *Erdős-Rényi networks*. In the asymptotic analysis of Erdős-Rényi networks, the density $p(n)$ plays a pivotal role in shaping the structural attributes of the network. Some critical cases are outlined as follows:

- When $p(n) = o(n^{-2})$, the Erdős-Rényi networks are empty almost surely (Erdős et al. 1960).
- When $p(n) = \mathcal{O}(n^{-(1+\epsilon)})$ for some $\epsilon > 0$, the expected in-degree and out-degree vanishes asymptotically. Such networks are called *very sparse* networks. They are probabilistically acyclic and fragmented.
- When $p(n) = \Theta(n^{-1})$, the expected in-degree and out-degree remain asymptotically bounded and positive. Such networks are called *critically sparse* networks. At this point, a phase transition occurs: as $p(n)$ increases from $\frac{1}{n} - \mathcal{O}(n^{-\frac{4}{3}})$ to $\frac{1}{n} + O(n^{-\frac{4}{3}})$, smaller components merge into a giant component comprising a positive fraction of nodes, and cycles begin to form (Janson et al. 1993).
- When $p(n) = \omega(\frac{\log n}{n})$, networks are called *dense* networks. These networks are highly likely to be connected, contain many cycles, and are asymptotically regular and balanced. Both in-degree and out-degree distributions concentrate around the mean value and converge asymptotically to a normal distribution.

It is important to note that dense networks are asymptotically regular and balanced. This underlying property aligns our numerical findings with the theoretical implications with regard to the imbalance level of networks. In Figure EC.8, we illustrate the distribution of in-degree and out-degree pairs for each node in the Erdős-Rényi networks with different densities. As evidenced by Figure EC.8, the distribution of in-degrees and out-degrees approaches a normal distribution as $p(n)$ increases. More importantly, this trend also shows that the in-degree and the out-degree of a specific node become close to each other, contributing to a more balanced network structure.

We further complement our analysis by singling out agents with low in-degrees, who, as illustrated by both numerical and theoretical results, exhibit poor performance in the FPA scheme. We specifically analyze the MAPE for nodes with $d = 1, 2$, and those with $d \geq 5$ for comparison, visualized in Figure EC.9, across various network sizes and densities. Key takeaways include the following: Agents with an in-degree of 1 exhibit larger MAPE compared to those with larger in-degrees, as shown in Figures 3a and 4a. These results reinforce our node-level theoretical analyses (Theorems 1 and 2). Additionally, for agents with the same

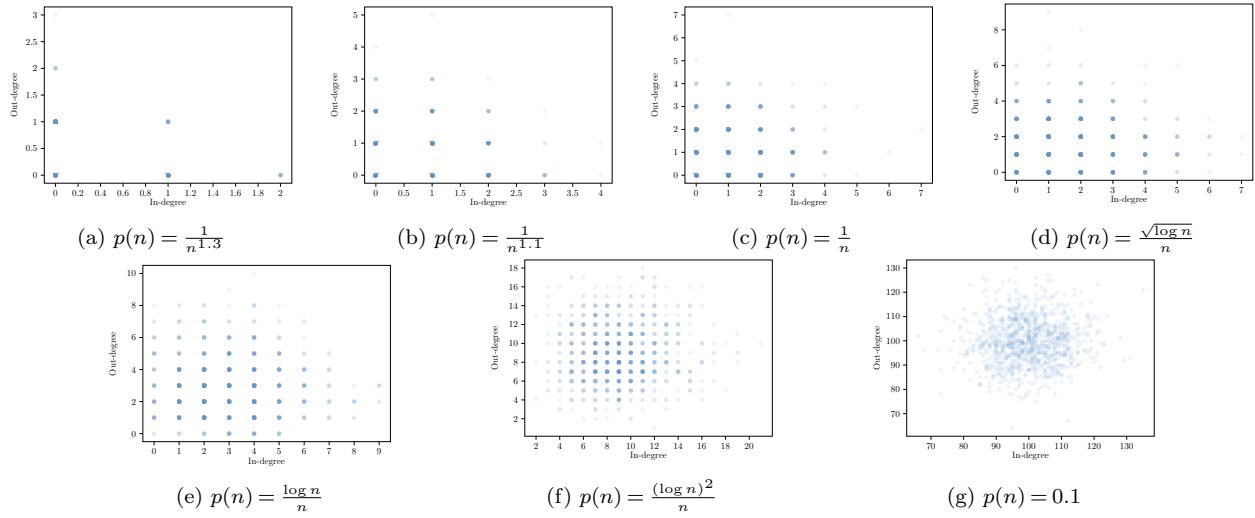


Figure EC.8 In-degree and out-degree distributions of Erdős-Rényi networks with different densities $p(n)$.

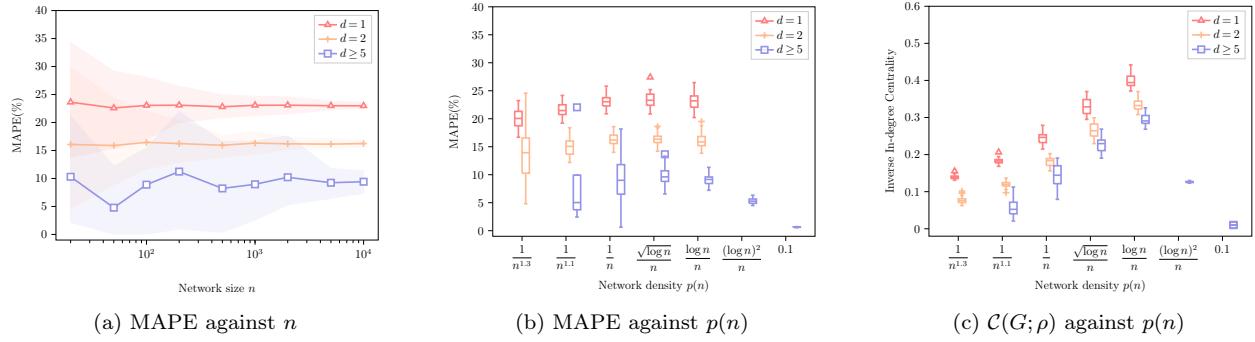


Figure EC.9 Degree-level FPA performance. In the left subfigure, $p(n)$ is fixed to be $\frac{1}{n}$ and the horizontal axis is in log scale. Shaded areas represent the 95% confidence interval. In the middle subfigure, n is fixed to be 1,000.

in-degree, we find that their MAPEs remain relatively stable as n increases, though there is a slight upward trend as $p(n)$ increases. This trend can be attributed to the influence of more distant neighbors, highlighting the importance of capturing the overall network structure and connectivity. Furthermore, we find that this structural information is well represented by the inverse in-degree centrality, as shown in Figure EC.9c. The patterns in the centrality measure closely match those of MAPE, and this alignment would be nearly perfect if we used the mean absolute error instead. However, we have omitted those results for brevity.

(ii) *Power-law networks*. To avoid notation confusion, in this part, we let d_i^{in} and d_i^{out} denote the in-degree and out-degree of node i . To construct a power-law network, we require both the in-degrees $d_1^{\text{in}}, d_2^{\text{in}}, \dots, d_n^{\text{in}}$ and the out-degrees $d_1^{\text{out}}, d_2^{\text{out}}, \dots, d_n^{\text{out}}$ are i.i.d. sampled from distribution $F_d(\cdot, n)$. One distinctive aspect of our experiments with power-law networks is the introduction of a pairwise correlation parameter to capture the imbalance level of the network. A valid correlated in-degree sequence and out-degree sequence can be generated using the following procedure:

- Sample i.i.d. in-degrees $d_1^{\text{in}}, d_2^{\text{in}}, \dots, d_n^{\text{in}}$ from the power-law distribution. Without loss of generality, assume this sequence is sorted in descending order.

- Sample i.i.d. random variables Z_1, Z_2, \dots, Z_n as follows: for each $i \in V$, $Z_i = 1$ with probability $|\theta|$, and $Z_i = 0$ with probability $1 - |\theta|$, where $\theta \in [-1, 1]$ is the parameter used to control the correlation. This parameter θ is not necessarily the correlation $\text{Cov}(d_i^{\text{in}}, d_i^{\text{out}})$.
- Define sets of nodes $I_0 : \{i : Z_i = 0, 1 \leq i \leq n\}$ and $I_1 : \{i : Z_i = 1, 1 \leq i \leq n\}$.
- If $\theta \geq 0$, set $d_i^{\text{out}} = d_i^{\text{in}}$ for $i \in I_1$ and set $\{d_i^{\text{out}} : i \in I_0\}$ by a random permutation of $\{d_i^{\text{in}} : i \in I_0\}$; If $\theta < 0$, set $d_i^{\text{out}} = d_{n-i+1}^{\text{in}}$ for $i \in I_0$ and set $\{d_i^{\text{out}} : i \in I_1\}$ by a random permutation of $\{d_{n-i+1}^{\text{in}} : i \in I_1\}$.
- Use a configuration model (Molloy and Reed 1995, Newman et al. 2001) to construct the directed random network with given in- and out-degree sequences.

Under this construction, when $\theta \geq 0$, the correlation is $\theta + O(n^{-1})$, so it asymptotically equals θ . When $\theta < 0$, the generated pairwise correlation may deviate from θ , and different values of θ yield similar degree sequences, as evidenced by Figure EC.10.

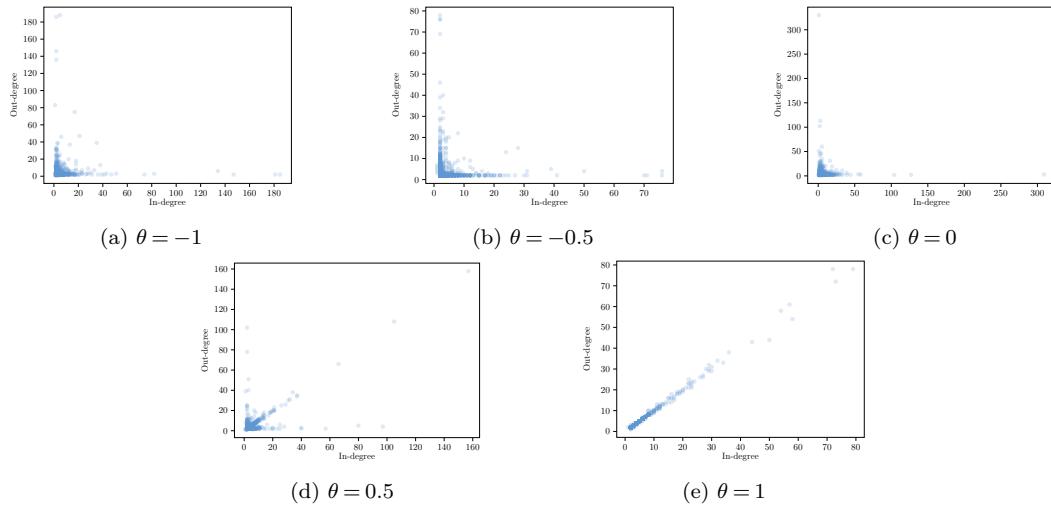


Figure EC.10 In-degree and out-degree distributions of power-law networks with different pairwise correlations θ .

Appendix E: Supporting Arguments for Section 6

E.1. Regret Analysis of Optimization Problem

We first analyze how the approximation error of the FPA scheme translates into the optimality gap of the optimization problem. Considering the general optimization problem (4) and its approximate formulation (11), we define the regret for a platform decision \mathbf{x} as the difference between the optimal objective value and the objective value under \mathbf{x} . Formally, the regret is given by:

$$\text{Regret}(\mathbf{x}) := g\left(\mathbf{q}^*(G, \mathbf{v}(\mathbf{x}^*), F_\epsilon(\cdot), \beta), \mathbf{x}^*\right) - g\left(\mathbf{q}^*(G, \mathbf{v}(\mathbf{x}), F_\epsilon(\cdot), \beta), \mathbf{x}\right), \quad (\text{EC.17})$$

where \mathbf{x}^* is the optimal solution derived from the original problem (4).

We next derive the regret bounds for the IM problem and the optimal pricing problem, respectively.

For the IM problem, the regret bound for the optimal solution S^{FPA} to (19) compared to the optimal solution to (18), S^* , is established in Proposition EC.1.

PROPOSITION EC.1 (Regret Bound for the IM Problem). *Under Assumptions 1 and 2, for any IM instance $(G, \mathbf{v}, F_\epsilon(\cdot), \beta)$, $\text{Regret}(S^{\text{FPA}}) \leq 2C_\rho \sqrt{n\|\mathcal{C}(G; \rho)\|_1}$.*

The approximation error of the FPA scheme directly translates into a decision error. All of our previous findings, including the refined bounds discussed in Section 4, extend to the approximate IM problem. For example, following (7), the worst-case regret bound in Proposition EC.1 can also be adjusted to an order of $\mathcal{O}(n/\sqrt{d_{\min}})$, which is sublinear in n when d_{\min} increases with rate $\Omega(1)$.

Proof of Proposition EC.1: The proof largely follows Corollary 2, where we bound the scaled ℓ_1 -norm of the FPA error. Therefore, the regret can be bounded by

$$\begin{aligned} \text{Regret}(S^{\text{FPA}}) &= \sum_{i \in V} q_i^*(S^*) - \sum_{i \in V} q_i^*(S^{\text{FPA}}) \\ &= \sum_{i \in V} q_i^*(S^*) - \sum_{i \in V} \mu_i^*(S^*) + \sum_{i \in V} \mu_i^*(S^*) - \sum_{i \in V} \mu_i^*(S^{\text{FPA}}) + \sum_{i \in V} \mu_i^*(S^{\text{FPA}}) - \sum_{i \in V} q_i^*(S^{\text{FPA}}) \\ &\leq \left\| \mathbf{q}^*(S^*) - \boldsymbol{\mu}^*(S^*) \right\|_1 + \left(\sum_{i \in V} \mu_i^*(S^*) - \sum_{i \in V} \mu_i^*(S^{\text{FPA}}) \right) + \left\| \mathbf{q}^*(S^{\text{FPA}}) - \boldsymbol{\mu}^*(S^{\text{FPA}}) \right\|_1 \\ &\leq \left\| \mathbf{q}^*(S^*) - \boldsymbol{\mu}^*(S^*) \right\|_1 + \left\| \mathbf{q}^*(S^{\text{FPA}}) - \boldsymbol{\mu}^*(S^{\text{FPA}}) \right\|_1 \\ &\leq 2C_\rho \sqrt{n\|\mathcal{C}(G; \rho)\|_1}, \end{aligned}$$

where the first inequality holds trivially, the second inequality follows from the optimality of S^{FPA} for approximate IM problem (19), and the third inequality follows from Corollary 2. \square

For the optimal pricing problem, we establish the regret bound for the optimal solution to (21), denoted by \mathbf{p}^{FPA} in Proposition EC.2.

PROPOSITION EC.2 (Regret Bound for the Pricing Problem). *Under Assumptions 1 and 2, for any pricing instance $(G, \mathbf{v}, F_\epsilon(\cdot), \beta)$, $\text{Regret}(\mathbf{p}^{\text{FPA}}) \leq p_{\max} C_\rho \sqrt{n\|\mathcal{C}(G; \rho)\|_1}$, where $p_{\max} := \max \left\{ \|\mathbf{p}^*\|_\infty, \|\mathbf{p}^{\text{FPA}}\|_\infty \right\}$.*

Proposition EC.2 establishes a regret bound similar to Proposition EC.1, except that the bound depends on p_{\max} , which can be seen as the maximal derivative of the objective function with regard to the adoption probability. In practice, the platform usually has a natural upper bound for prices (e.g., the price under which no agent will make a purchase, regardless of their neighbors' adoption decisions). Therefore, p_{\max} can be bounded by a constant.

Proof of Proposition EC.2: The proof aligns with Proposition EC.1. Therefore, the regret can be bounded by

$$\begin{aligned} \text{Regret}(\mathbf{p}^{\text{FPA}}) &= \mathbf{q}^*(\mathbf{p}^*)^\top \mathbf{W} \mathbf{p}^* - \mathbf{q}^*(\mathbf{p}^{\text{FPA}})^\top \mathbf{W} \mathbf{p}^{\text{FPA}} \\ &= \mathbf{q}^*(\mathbf{p}^*)^\top \mathbf{W} \mathbf{p}^* - \boldsymbol{\mu}^*(\mathbf{p}^*)^\top \mathbf{W} \mathbf{p}^* + \boldsymbol{\mu}^*(\mathbf{p}^*)^\top \mathbf{W} \mathbf{p}^* - \boldsymbol{\mu}^*(\mathbf{p}^{\text{FPA}})^\top \mathbf{W} \mathbf{p}^{\text{FPA}} \\ &\quad + \boldsymbol{\mu}^*(\mathbf{p}^{\text{FPA}})^\top \mathbf{W} \mathbf{p}^{\text{FPA}} - \mathbf{q}^*(\mathbf{p}^{\text{FPA}})^\top \mathbf{W} \mathbf{p}^{\text{FPA}} \\ &\leq \left\| (\mathbf{q}^*(\mathbf{p}^*) - \boldsymbol{\mu}^*(\mathbf{p}^*))^\top \mathbf{W} \mathbf{p}^* \right\|_1 + \left(\boldsymbol{\mu}^*(\mathbf{p}^*)^\top \mathbf{W} \mathbf{p}^* - \boldsymbol{\mu}^*(\mathbf{p}^{\text{FPA}})^\top \mathbf{W} \mathbf{p}^{\text{FPA}} \right) \\ &\quad + \left\| (\mathbf{q}^*(\mathbf{p}^{\text{FPA}}) - \boldsymbol{\mu}^*(\mathbf{p}^{\text{FPA}}))^\top \mathbf{W} \mathbf{p}^{\text{FPA}} \right\|_1 \end{aligned}$$

$$\begin{aligned}
&\leq \left\| (\mathbf{q}^*(\mathbf{p}^*) - \boldsymbol{\mu}^*(\mathbf{p}^*))^\top \mathbf{W} \mathbf{p}^* \right\|_1 + \left\| (\mathbf{q}^*(\mathbf{p}^{\text{FPA}}) - \boldsymbol{\mu}^*(\mathbf{p}^{\text{FPA}}))^\top \mathbf{W} \mathbf{p}^{\text{FPA}} \right\|_1 \\
&\leq \left\| \mathbf{q}^*(\mathbf{p}^*) - \boldsymbol{\mu}^*(\mathbf{p}^*) \right\|_1 \left\| \mathbf{W} \mathbf{p}^* \right\|_\infty + \left\| \mathbf{q}^*(\mathbf{p}^{\text{FPA}}) - \boldsymbol{\mu}^*(\mathbf{p}^{\text{FPA}}) \right\|_1 \left\| \mathbf{W} \mathbf{p}^{\text{FPA}} \right\|_\infty \\
&\leq 2C_\rho \max \left\{ \left\| \mathbf{p}^* \right\|_\infty, \left\| \mathbf{p}^{\text{FPA}} \right\|_\infty \right\} \sqrt{n \|\mathcal{C}(G; \rho)\|_1},
\end{aligned}$$

where the first inequality follows trivially, the second inequality follows since the optimality of \mathbf{p}^{FPA} for approximate pricing problem (21), the third inequality follows from Corollary 2, and the last one follows the properties of matrix operator norms. \square

E.2. Supporting Arguments for Section 6.1

In the subsection, we first discuss the applicability of Assumption 4.

Instances that satisfy Assumption 4. As we mentioned immediately after the assumption, the non-progressive LT model is a specific instance that meets this assumption. Recall we can recover the LT model by setting $v_i = -0.5$ and $\epsilon_i(t) \sim \mathcal{U}(-0.5, 0.5)$ for all $i \in V$ and $t \geq 1$. Therefore, for any $\beta > 0$, CDF F_ϵ can be expressed as $F_\epsilon(x) = \mathbb{1}\{x \geq -0.5\} \cdot (x + 0.5)$ on range $[0.5 - \beta, 0.5]$, which is convex. Additionally, some other diffusion instances related to common utility models can also meet Assumption 4. Some examples are: (i) Linear probability model: $v_i \geq -c$ and $\epsilon_i(t) \sim \mathcal{U}(-c, c)$ for all $i \in V$, $t \geq 1$. (ii) Logit model: $v_i \geq 0$ and $\epsilon_i(t) \sim \text{Logistic}(0, s)$ and for all $i \in V$, $t \geq 1$. (iii) Probit model: $v_i \geq 0$ and $\epsilon_i(t) \sim \mathcal{N}(0, s)$ and for all $i \in V$, $t \geq 1$. For many general distributions, the convexity assumption essentially requires the intrinsic values to be appropriately lower bounded.

Then, we show the proof of the submodularity for the approximate IM objective.

Proof of Theorem 4: Consider two seed set $S_1 \subseteq S_2 \subseteq V$ and an additional user $w \in V \setminus S_2$, it is sufficient to show that $\boldsymbol{\mu}^*(S_2 + \{w\}) - \boldsymbol{\mu}^*(S_2) \leq \boldsymbol{\mu}^*(S_1 + \{w\}) - \boldsymbol{\mu}^*(S_1)$.

We consider constraints (19b) and (19c) as the dynamic system, that is, $\boldsymbol{\mu}(t) = \mathbf{h}(\boldsymbol{\mu}(t-1))$. We can notice that, for different seed sets, the transition function \mathbf{h} is not the same. However, for all the users that are not selected as seed users, the transition function for the corresponding element is the same. With a little abuse of notation, in the following proof, we use \mathbf{h} to denote the transition function for all users in $V \setminus (S_2 + \{w\})$.

We want to show that at any time step $t \geq 1$, the inequality $\boldsymbol{\mu}(S_2 + \{w\}, t) - \boldsymbol{\mu}(S_2, t) \leq \boldsymbol{\mu}(S_1 + \{w\}, t) - \boldsymbol{\mu}(S_1, t)$ always holds. For user $i \in S_2$, $\mu_i(S_2 + \{w\}, t) - \mu_i(S_2, t) = 0 \leq \mu_i(S_1 + \{w\}, t) - \mu_i(S_1, t)$. For user w , $\mu_w(S_2 + \{w\}, t) - \mu_w(S_2, t) = 1 - \mu_w(S_2, t) \leq 1 - \mu_w(S_1, t) \leq \mu_w(S_1 + \{w\}, t) - \mu_w(S_1, t)$. The above two inequalities hold because of Proposition 2(i). For all the other users in $V \setminus (S_2 \cup \{w\})$, we show by induction.

Base case $t = 0$: First of all, $\boldsymbol{\mu}(S, 0) = \mathbf{0}$ for all $S \subseteq S_2 \cup \{w\}$ by definition. Therefore,

$$\boldsymbol{\mu}(S_2 + \{w\}, 0) - \boldsymbol{\mu}(S_2, 0) = \boldsymbol{\mu}(S_1 + \{w\}, 0) - \boldsymbol{\mu}(S_1, 0).$$

Assume $t = s$: The induction hypothesis holds such that

$$\boldsymbol{\mu}(S_2 + \{w\}, s) - \boldsymbol{\mu}(S_2, s) \leq \boldsymbol{\mu}(S_1 + \{w\}, s) - \boldsymbol{\mu}(S_1, s).$$

Show $t = s+1$: We have

$$\begin{aligned}
\boldsymbol{\mu}(S_2 + \{w\}, s+1) - \boldsymbol{\mu}(S_2, s+1) &= \mathbf{h}(\boldsymbol{\mu}(S_2 + \{w\}, s)) - \mathbf{h}(\boldsymbol{\mu}(S_2, s)) \\
&\leq \mathbf{h}(\boldsymbol{\mu}(S_2, s) + \boldsymbol{\mu}(S_1 + \{w\}, s) - \boldsymbol{\mu}(S_1, s)) - \mathbf{h}(\boldsymbol{\mu}(S_2, s)) \\
&\leq \mathbf{h}(\boldsymbol{\mu}(S_1 + \{w\}, s)) - \mathbf{h}(\boldsymbol{\mu}(S_1, s)) \\
&= \boldsymbol{\mu}(S_1 + \{w\}, s+1) - \boldsymbol{\mu}(S_1, s+1),
\end{aligned}$$

where the first inequality comes from Proposition 2(i) and the second inequality comes from Assumption 4.

When t tends to infinity, we get the fixed-point solution $\mu^*(S_2 + \{w\}) - \mu^*(S_2) \leq \mu^*(S_1 + \{w\}) - \mu^*(S_1)$, and hence the submodularity is proved. \square

E.3. Experiments on IM Problem

In the experiments, we consider two scenarios, one satisfies Assumption 4 and thus leads to a submodular influence function, while the other does not. For both scenarios, we assume that the intrinsic value $v_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{U}(-4, 0)$ and $\beta = 3.5$. In addition, we assume the random noise to be $\epsilon_i(t) \stackrel{\text{i.i.d.}}{\sim} \mathcal{U}(-4, 4)$ in the submodular case, while $\epsilon_i(t) \stackrel{\text{i.i.d.}}{\sim} \text{Logistic}(0, 1)$ in the nonsubmodular case.

The well-known greedy framework selects one user at each iteration which leads to the largest total adoptions. We refer to the algorithm that embeds the FPA solution into this greedy framework for the total influence evaluation as the **greedy-FPA** algorithm. We randomly generate some small network instances to illustrate that **greedy-FPA** can find a near-optimal solution. Although there is no theoretical guarantee for the nonsubmodular case, it is interesting to observe from the results that the FPA solutions are still of good quality. For either scenario, we generate 100 diffusion instances with random graph $G(15, 0.5)$ and set the number of seed users to 5. We enumerate all the subsets to find the optimal seed set and evaluate the diffusion influence using Naïve ABS method. In Table EC.2, we show the numerical results of the **greedy-FPA** algorithm.

Table EC.2 Numerical results of greedy-FPA algorithm for IM problem.

Scenario	Percentage of instances where the optimal seeding is recovered	Optimality Gap (%)	
		Mean	Max
Submodular	91	0.0194	0.4704
Non-submodular	84	0.0685	1.7444

We notice that in both the submodular and nonsubmodular scenarios, the **greedy-FPA** algorithm can generate a near-optimal IM solution and even uncover the exact optimal solution for a large portion of instances. Meanwhile, the **greedy-FPA** algorithm has a slightly better performance in the submodular case than in the nonsubmodular case but even in the nonsubmodular problem instances, it remains quite practical.

Furthermore, we choose a real-world network—*Caltech36* as introduced in Section 5.3 and compare the performance of **greedy-FPA** with the traditional IM heuristics. Recall that the instance includes 765 agents with an average number of neighbors of 43. We define several benchmark strategies as follows. The **DEG** and **EIG** schemes are motivated by the important role of the centrality measures in diffusion discussed in the network economics literature (e.g., Ballester et al. 2006, Jackson 2010). We include them for completeness, but as substantiated in the numerical experiments, by overlooking the idiosyncratic features of the agents, these schemes are dominated by the FPA-based heuristic.

Benchmarks:

- *Greedy and Naïve ABS* (greedy-ABS): This is the classical algorithm used for the IM problem. The Naïve ABS is embedded into a greedy framework for influence evaluation. The length of the Naïve ABS run is set to 100,000 after the warm-up period.
- *Greedy and the FPA solution* (greedy-FPA): This is our proposed algorithm. We embed the FPA solution into a greedy framework for an influence evaluation.
- *Greedy and low-resolution ABS* (greedy-low-ABS): The Naïve ABS is embedded into a greedy framework for an influence evaluation. The length of the Naïve ABS run is set to 50 so that the runtime is at the same scale as that of the FPA scheme.
- *Degree centrality* (DEG): Set K users with the largest degrees to be seed users.
- *Eigenvector centrality* (EIG): Set K users with the largest eigenvector centralities to be seed users.
- *Model misspecification without network effect* (MM): This benchmark considers the misspecified model that ignores the network effect in the IM problem. This is the same as setting K users with the smallest intrinsic value to be seed users.
- *Random* (RAN): Randomly select K users to be seed users.

Figure EC.11 demonstrates the relative loss of the expected limiting adoptions compared with greedy-ABS against the number of seed users. Similarly, we also consider both the submodular and nonsubmodular cases.

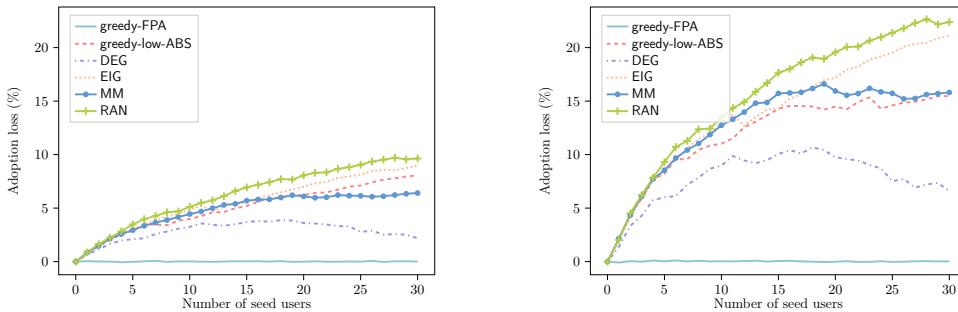


Figure EC.11 Performance of different IM algorithms. Left: Submodular case; Right: Non-submodular case.

When the number of seed users increases from 0 to 30, the difficulty of the IM problem increases since the number of feasible solutions also increases. We observe that regardless of the number of seed users, the performance of greedy-FPA matches that of greedy-ABS nearly perfectly. It significantly outperforms all the other benchmarks. This is no surprise to us, as it is driven by the high accuracy of the FPA scheme. In particular, we also notice that the performance of the greedy framework with the Naïve ABS method degrades drastically when the simulation length of the Naïve ABS procedure is small. Compared with greedy-low-ABS, greedy-FPA achieves an improvement of 8.90% and 18.42% when $K = 30$ in the submodular and nonsubmodular cases. In short, we conclude that, by offering a significant efficiency gain, greedy-FPA outperforms greedy-ABS in solving the IM problem.

E.4. Supporting Arguments for Section 6.2

In this subsection, we first focus on the proof of the pricing problem in the adoption probability space.

Proof of Theorem 5: Let $\pi(\boldsymbol{\mu}) = \sum_{i \in V} \left(v_i + \beta \sum_{j \in \mathcal{N}_i} \frac{\mu_j}{d_i} + \ln \frac{1-\mu_i}{\mu_i} \right) \mu_i$. The Hessian matrix of $\pi(\boldsymbol{\mu})$ can be derived as

$$\frac{\partial^2 \pi}{\partial \mu_i^2} = -\frac{1}{\gamma} \frac{1}{\mu_i(\mu_i-1)^2} \quad \text{and} \quad \frac{\partial^2 \pi}{\partial \mu_i \partial \mu_j} = \frac{1}{\gamma} \mathbb{1}\{j \in \mathcal{N}_i\} \frac{\beta}{d_i} + \frac{1}{\gamma} \mathbb{1}\{i \in \mathcal{N}_j\} \frac{\beta}{d_j}.$$

For the diagonal elements of the Hessian matrix H_π , we can have $-1/[\mu(\mu-1)^2] \leq -6.75$ holds for any $x \in [0, 1]$. The inequality is tight when $\mu = 1/3$. For the nondiagonal elements of the Hessian matrix \mathbf{H}_π , we can find them related to the structure of network $G(V, E)$.

Therefore, we can have the Hessian matrix to be

$$\mathbf{H}_\pi = \frac{1}{\gamma} \text{diag} \left(\left\{ -\frac{1}{\mu_i(\mu_i-1)^2} \right\}_{i \in V} \right) + \frac{\beta}{\gamma} (\tilde{\mathbf{A}} + \tilde{\mathbf{A}}^\top) \preceq \frac{1}{\gamma} [-6.75I + \beta (\tilde{\mathbf{A}} + \tilde{\mathbf{A}}^\top)]$$

Therefore, when $\beta \leq \frac{6.25}{\lambda_{\max}(\tilde{\mathbf{A}} + \tilde{\mathbf{A}}^\top)}$,

$$\lambda_{\max} (-6.75I + \beta (\tilde{\mathbf{A}} + \tilde{\mathbf{A}}^\top)) \leq 0,$$

which implies that \mathbf{H}_π is negative semi-definite. The equality holds when $\beta = \frac{6.25}{\lambda_{\max}(\tilde{\mathbf{A}} + \tilde{\mathbf{A}}^\top)}$.

By Gershgorin circle theorem, we can bound the eigenvalues of $\tilde{\mathbf{A}}$ by $-1 \leq \lambda(\tilde{\mathbf{A}}) \leq 1$. Since 1 is one of the eigenvalues of $\tilde{\mathbf{A}}$, we can have $\lambda_{\max}(\tilde{\mathbf{A}}) = 1$. Therefore, when $\tilde{\mathbf{A}}$ is a symmetric matrix, $\frac{6.25}{\lambda_{\max}(\tilde{\mathbf{A}} + \tilde{\mathbf{A}}^\top)} = 3.375$.

□

The proof for the concavity of dynamic pricing objective largely mirrors that of the static pricing problem.

Proof of Theorem 6: Let $\pi(\boldsymbol{\mu}) = \sum_{i \in V} \left(v_i + \beta \sum_{j \in \mathcal{N}_i} \frac{\mu_j(t-1)}{d_i} + \ln \frac{1-\mu_i(t)}{\mu_i(t)} \right) \mu_i(t)$. The Hessian matrix of $\pi(\boldsymbol{\mu})$ can be derived as

$$\frac{\partial^2 \pi}{\partial \mu_i(t)^2} = -\frac{1}{\gamma} \frac{1}{\mu_i(t)(\mu_i(t)-1)^2}, \quad \frac{\partial^2 \pi}{\partial \mu_i(t) \partial \mu_j(t-1)} = \frac{1}{\gamma} \mathbb{1}\{j \in \mathcal{N}_i\} \frac{\beta}{d_i} \quad \text{and} \quad \frac{\partial^2 \pi}{\partial \mu_i(t-1) \partial \mu_j(t)} = \frac{1}{\gamma} \mathbb{1}\{i \in \mathcal{N}_j\} \frac{\beta}{d_j}$$

For the diagonal elements of the Hessian matrix H_π , we can have $-1/[\mu(\mu-1)^2] \leq -6.75$ holds for any $x \in [0, 1]$. The inequality is tight when $\mu = 1/3$. For the nondiagonal elements of the Hessian matrix \mathbf{H}_π , we can find them related to the network structure. Let us define the following block matrix:

$$\mathbf{M} := \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \tilde{\mathbf{A}} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{A}} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & & & & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \tilde{\mathbf{A}} & \mathbf{0} \end{pmatrix}$$

Therefore, we can represent the Hessian matrix as the following block matrix

$$\begin{aligned} & \frac{1}{\gamma} \begin{pmatrix} \text{diag} \left(\left\{ -\frac{1}{\mu_i(\mu_i-1)^2} \right\}_{i \in V} \right) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \text{diag} \left(\left\{ -\frac{1}{\mu_i(\mu_i-1)^2} \right\}_{i \in V} \right) & \cdots & \mathbf{0} \\ \vdots & & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \text{diag} \left(\left\{ -\frac{1}{\mu_i(\mu_i-1)^2} \right\}_{i \in V} \right) \end{pmatrix} + \frac{\beta}{\gamma} (\mathbf{M} + \mathbf{M}^\top) \\ & \preceq \frac{1}{\gamma} [-6.75I + \beta (\mathbf{M} + \mathbf{M}^\top)]. \end{aligned}$$

For any block vector $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_t)$, we have

$$\begin{aligned}\mathbf{x}^\top (\mathbf{M} + \mathbf{M}^\top) \mathbf{x} &= \sum_{t=1}^{T-1} \mathbf{x}_t^\top \tilde{\mathbf{A}}^\top \mathbf{x}_{t+1} + \mathbf{x}_{t+1}^\top \tilde{\mathbf{A}} \mathbf{x}_t = \sum_{t=1}^{T-1} \mathbf{x}_t^\top (\tilde{\mathbf{A}} + \tilde{\mathbf{A}}^\top) \mathbf{x}_{t+1} \\ &\leq \left\| \tilde{\mathbf{A}} + \tilde{\mathbf{A}}^\top \right\|_2 \cdot \sum_{t=1}^{T-1} \|\mathbf{x}_t\|_2 \|\mathbf{x}_{t+1}\|_2,\end{aligned}$$

where the inequality follows from the definition of spectral norm.

Define the scalar vector $\mathbf{s} = (\|\mathbf{x}_1\|_2, \dots, \|\mathbf{x}_t\|_2)$, we have

$$\mathbf{s}^\top \tilde{\mathbf{D}} \mathbf{s} = 2 \sum_{\tau=1}^{t-1} s_t s_{t+1},$$

where $\tilde{\mathbf{D}}$ is the $t \times t$ tridiagonal with ones on the super-/sub-diagonals and zeros on the diagonal. Hence,

$$\sum_{t=1}^{T-1} s_t s_{t+1} \leq \frac{1}{2} \lambda_{\max}(\tilde{\mathbf{D}}) \|\mathbf{s}\|_2^2 = \cos\left(\frac{\pi}{T+1}\right) \|\mathbf{x}\|_2^2.$$

Taking the supremum over $\|\mathbf{x}\|_2 = 1$ gives us

$$\lambda_{\max}(\mathbf{M} + \mathbf{M}^\top) \leq \|\mathbf{A} + \mathbf{A}^\top\|_2 \cos\left(\frac{\pi}{T+1}\right).$$

Therefore, when $\beta \leq \frac{6.25}{\lambda_{\max}(\tilde{\mathbf{A}} + \tilde{\mathbf{A}}^\top) \cos(\frac{\pi}{T+1})}$,

$$\lambda_{\max}(-6.75I + \beta(\tilde{\mathbf{A}} + \tilde{\mathbf{A}}^\top)) \leq 0,$$

which implies that \mathbf{H}_π is negative semi-definite. When $T \geq 1$, we also have $\cos\left(\frac{\pi}{T+1}\right) \leq 1$.

Since $\lambda_{\max}(\tilde{\mathbf{A}}) = 1$, when $\tilde{\mathbf{A}}$ is a symmetric matrix, $\frac{6.25}{\lambda_{\max}(\tilde{\mathbf{A}} + \tilde{\mathbf{A}}^\top) \cos(\frac{\pi}{T+1})} = \frac{3.375}{\cos(\frac{\pi}{T+1})}$. \square

\square

In the following, we illustrate the procedure of gradient descent for the pricing problem in the price space.

Gradient descent and approximate gradient descent. By taking the derivative on both sides of the fixed-point equation, we get

$$\frac{d\boldsymbol{\mu}(\mathbf{p})}{d\mathbf{p}} = \frac{\partial \mathbf{h}(\mathbf{p}, \boldsymbol{\mu}(\mathbf{p}))}{\partial \mathbf{p}} + \frac{d\boldsymbol{\mu}(\mathbf{p})}{d\mathbf{p}} \cdot \frac{\partial \mathbf{h}(\mathbf{p}, \boldsymbol{\mu}(\mathbf{p}))}{\partial \boldsymbol{\mu}(\mathbf{p})}.$$

By rearranging the terms, we obtain

$$\frac{d\boldsymbol{\mu}(\mathbf{p})}{d\mathbf{p}} \cdot \left(\mathbf{I} - \frac{\partial \mathbf{h}(\mathbf{p}, \boldsymbol{\mu}(\mathbf{p}))}{\partial \boldsymbol{\mu}(\mathbf{p})} \right) = \frac{\partial \mathbf{h}(\mathbf{p}, \boldsymbol{\mu}(\mathbf{p}))}{\partial \mathbf{p}}.$$

Matrix $(\mathbf{I} - \partial \mathbf{h}(\mathbf{p}, \boldsymbol{\mu}(\mathbf{p}))/\partial \boldsymbol{\mu}(\mathbf{p}))$ is guaranteed to be invertible. The reason is that, by Proposition 2, we know that \mathbf{h} is a contraction mapping and $\|\partial \mathbf{h}(\mathbf{p}, \boldsymbol{\mu}(\mathbf{p}))/\partial \boldsymbol{\mu}(\mathbf{p})\|_\infty < 1$.

With an eye toward implementation, we also notice that (25) requires computing the inverse of an n -by- n matrix. When the network is large and dense, this calculation becomes intimidating. Fortunately, we know that $\|\partial \mathbf{h}(\mathbf{p}, \boldsymbol{\mu}(\mathbf{p}))/\partial \boldsymbol{\mu}(\mathbf{p})\|_\infty < 1$, and hence, the spectral radius of $\partial \mathbf{h}(\mathbf{p}, \boldsymbol{\mu}(\mathbf{p}))/\partial \boldsymbol{\mu}(\mathbf{p})$ is smaller than 1. Consequently, we can expand the inverse as the Neumann series, which is similar to the centrality measure,

$$\left(\mathbf{I} - \frac{\partial \mathbf{h}(\mathbf{p}, \boldsymbol{\mu}(\mathbf{p}))}{\partial \boldsymbol{\mu}(\mathbf{p})} \right)^{-1} = \mathbf{I} + \sum_{\ell=1}^{\infty} \left(\frac{\partial \mathbf{h}(\mathbf{p}, \boldsymbol{\mu}(\mathbf{p}))}{\partial \boldsymbol{\mu}(\mathbf{p})} \right)^\ell.$$

This leads to the following k -th order approximate gradient:

$$\frac{d\Pi(\mathbf{p})}{d\mathbf{p}} \approx \tilde{\mathbf{G}}_k(\mathbf{p}) = \frac{\partial \mathbf{h}(\mathbf{p}, \boldsymbol{\mu}(\mathbf{p}))}{\partial \mathbf{p}} \cdot \left(\mathbf{I} + \sum_{\ell=1}^k \left(\frac{\partial \mathbf{h}(\mathbf{p}, \boldsymbol{\mu}(\mathbf{p}))}{\partial \boldsymbol{\mu}(\mathbf{p})} \right)^\ell \right) \cdot \mathbf{W} \cdot \mathbf{p} + \mathbf{W}^\top \cdot \boldsymbol{\mu}(\mathbf{p}) \quad (\text{EC.18})$$

for the pricing problem. We expect such an easy-to-compute approximate gradient to lead to a significant efficiency gain, as is usually the case in the literature regarding approximate gradient descent (Ruder 2016). Previous works have applied similar low-order approximations for network effects for different purposes (e.g., see Candogan et al. 2012, Zeng et al. 2023). In subsequent numerical experiments, we find that $k = 2$ works very well in practice, leading to near-optimal solutions very quickly.

E.5. Experiments on Pricing Problems

In the experiments for pricing problems on a social network, the first issue is that the optimal pricing problem under the original diffusion model seems impossible to derive. We test over different randomly generated instances and find that the profits calculated via simulation and the FPA scheme are quite close, with a percentage error almost uniformly bounded by 0.5% in our experiment.

In order to check FPA performance about the total profit when price is considered, we test over three groups of instances. By fixing the expected number of neighbors to be 10, we generate diffusion instances with random graphs $G(20, 0.5)$, $G(100, 0.1)$, $G(1000, 0.01)$. For each instance, the agent is associated with an intrinsic value i.i.d. sampled from $\mathcal{U}(0, 4)$ and an offered price i.i.d. sampled from $\mathcal{U}(0, 4)$. We set $\beta = 3$ and $\gamma = 1$. In Figure EC.12, we show the distribution of profit difference among all diffusion instances. We notice

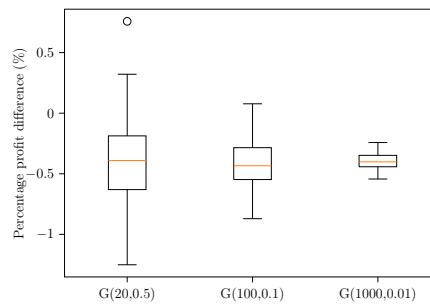


Figure EC.12 Profit difference between the ABS result and the FPA solution.

that the absolute profit difference is small. Furthermore, as the network becomes larger, the performance gap becomes more concentrated. In conclusion, we can consider the FPA scheme to be able to achieve almost the same performance as the simulation methods in the pricing problem.

Hereafter, we compare the pricing scheme under the FPA scheme as default. We assume $\epsilon_i(t) \stackrel{\text{i.i.d.}}{\sim} \text{Logistic}(0, 1)$, which follows the theoretical analysis in Section 6.2.

E.5.1. Static pricing problem. In this section, we consider the static pricing problem on a social network, considering limiting user behavior. We study two extreme scenarios, the *perfect price discrimination* case, where each consumer is offered a personal price, and the *public price* case, where all consumers receive the same price.

In the perfect price discrimination case, we test three different algorithms. The first algorithm is the gradient descent method in the adoption probability space (`grad-PROB`). With a network effect intensity that satisfies Theorem 5, `grad-PROB` can find the global optimal solution. The second algorithm is the gradient method in the price space (`grad-PRICE`). The third algorithm considers the pricing problem without network diffusion, that is, the price is determined according to the standard logit model. We still refer to it as the model misspecification (`MM`) scheme.

For the public price case, we also test three different algorithms. However, in this case, the pricing problem cannot be considered in the adoption probability space. Instead, we use a grid search (`GS`) to find an upper-bound solution to the problem. Specifically, we divide the price into grids of tolerance ξ . For each price p , we upper bound the profit with $(p - \xi) \cdot \sum_{i \in V} \mu_i(p)$. The other two algorithms, `grad-PRICE` and `MM`, as discussed above, are applied here.

For both scenarios, we test on a real-world network—*Amherst41* as introduced in Section 5.3. For each diffusion instance, we set the price sensitivity as $\gamma = 0.1$, and the intrinsic value $v_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{U}(0, 4)$. In Figure EC.13, we plot the realized profit of three algorithms against different values of the network effect intensity β . Furthermore, as remarked before, algorithm `grad-PRICE` involves the derivation of the gradient as (25), which requires calculating the inverse of a $n \times n$ matrix. We resort to the second-order approximate gradient, $\tilde{G}_2(\mathbf{p})$, as given in (EC.18).

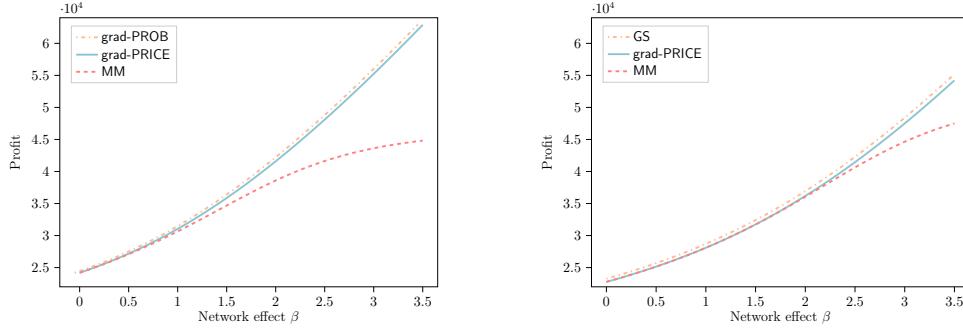


Figure EC.13 Realized profit versus network effect intensity. **Left:** Perfect price discrimination; **Right:** Public price (The curve of `grad-PROB` coincides with `grad-PRICE` in the left subfigure. In order to make them identifiable in the figures, we shift the `grad-PROB` to the left by 0.05.)

We offer several observations from this figure. First, `grad-PRICE` obtains a near-optimal solution in the case of price discrimination. This hints that we can use `grad-PRICE` to gain high-quality results in the general pricing setting when `grad-PROB` is not applicable. Second, there is a significant performance degradation of `MM` when the network effect is large. When $\beta = 3$, the relative profit loss reaches 21.16% and 8.30% if the network effect is ignored, respectively. When β increases to over 3.375 that violates the condition in 5, we can see that both algorithms `grad-PROB` and `grad-PRICE` still perform exceptionally. Third, comparing these two cases, we find that pricing discrimination can significantly increase the total profit, especially when the network effect intensity is large.

Furthermore, we compare the performance of these algorithms in more instances in terms of their execution time and the quality of solutions. We assume parameters $(n, p(n), \beta) \in \{100, 1000\} \times \{\frac{1}{n}, \frac{\log n}{n}, 0.1\} \times \{2, 3.5\}$. The numerical results for the two cases are shown in Tables EC.3 and EC.4. The grad-PRICE approach derives high-quality solutions in both cases. We notice that in the perfect price discrimination case, the run time for grad-PROB is less than grad-PRICE, although the margin is not too large. The profit difference between these two approaches is quite small, uniformly smaller than 0.25%. For the public price case, we set the tolerance of the grid search to be 0.5 within the range $[0, 100]$. grad-PRICE runs much faster than the grid search with a performance loss of up to 2%. The performance of MM remains poor across the two cases in this experiment, suggesting that the loss from ignoring network effects can be detrimental. Moreover, in the case when $\beta = 3.5$, we notice that the condition in 5 is violated, but practically, our algorithms provide commendable results. In summary, our main message through the numerical experiments is twofold. First, it is important to incorporate the network effect into operational problems. The gain from doing so can be significant. Second, we advocate grad-PRICE as a practical method for price optimization. With our approximate gradient expression tailored to the network setting as in (EC.18), grad-PRICE becomes a competitive price optimization technique. It can be efficiently implemented in various practical scenarios to find high-quality price solutions.

Table EC.3 Numerical results of pricing problem for randomly generated instances (perfect price discrimination)

Parameters ($n, p(n), \beta$)	grad-PROB		grad-PRICE			MM			
	time (s)	time (s)	profit loss (%)			time (s)	profit loss (%)		
			min	mean	max		min	mean	max
(100,1/n,2)	0.067	0.160	0.007	0.025	0.169	0.002	5.887	7.118	8.070
(100,log n/n,2)	0.056	0.154	0.026	0.038	0.053	0.002	5.717	6.670	7.759
(100,0.1,2)	0.044	0.128	0.031	0.040	0.058	0.002	5.238	6.199	6.990
(100,1/n,3.5)	0.034	0.090	0.247	0.023	0.842	0.001	19.824	22.625	25.161
(100,log n/n,3.5)	0.072	0.334	0.061	0.084	0.177	0.002	27.406	28.160	28.807
(100,0.1,3.5)	0.035	0.090	0.039	0.197	0.802	0.001	19.550	22.800	24.965
(1,000,1/n,2)	0.785	2.422	0.010	0.031	0.095	0.100	6.769	7.037	7.358
(1,000,log n/n,2)	0.576	1.986	0.035	0.039	0.043	0.099	6.139	6.390	6.697
(1,000,0.1,2)	0.696	1.917	0.037	0.041	0.047	0.099	5.376	5.878	6.251
(1,000,1/n,3.5)	0.616	1.658	0.114	0.233	0.429	0.080	21.794	22.798	23.591
(1,000,log n/n,3.5)	0.673	4.648	0.079	0.086	0.095	0.104	27.747	28.050	28.377
(1,000,0.1,3.5)	0.107	1.657	0.088	0.097	0.110	0.075	27.273	27.666	28.071

E.5.2. Dynamic pricing problem. In this section, we investigate the outcomes of applying dynamic pricing, as opposed to deriving the static optimal price once equilibrium is reached, under the perfect price discrimination case.

We evaluate three distinct pricing schemes: static, dynamic, and myopic. The static scheme utilizes the optimal static price derived in Section 6.2.1, assuming limiting consumer behavior. In contrast, the Dynamic scheme incorporates transient consumer behavior over time as in Section 6.2.2. Additionally, we introduce the Myopic benchmark, where the optimal price is recalculated at each time step based on the adoption levels

Table EC.4 Numerical results of pricing problem for randomly generated instances (public price)

Parameters ($n, p(n), \beta$)	GS		grad-PRICE			MM			
	time (s)	time (s)	profit loss (%)			time (s)	profit loss (%)		
			min	mean	max		min	mean	max
(100,1/n,2)	0.893	0.084	1.882	2.050	2.276	0.001	2.048	2.203	2.539
(100,log n/n,2)	1.364	0.087	1.882	2.226	0.013	0.001	2.085	2.411	3.051
(100,0.1,2)	1.075	0.073	1.874	2.038	2.260	0.001	2.088	2.441	3.036
(100,1/n,3.5)	1.065	0.083	1.721	1.870	2.029	0.001	2.608	4.216	6.380
(100,log n/n,3.5)	1.866	0.130	1.577	1.766	2.088	0.001	9.361	13.821	18.284
(100,0.1,3.5)	1.454	0.111	1.558	1.725	1.944	0.001	11.177	14.465	18.720
(1,000,1/n,2)	10.376	0.942	1.991	2.058	2.018	0.002	2.072	2.145	2.245
(1,000,log n/n,2)	12.891	0.916	1.981	2.036	2.100	0.002	2.235	2.402	2.587
(1,000,0.1,2)	13.057	0.975	1.971	2.027	2.093	0.002	2.247	2.408	2.594
(1,000,1/n,3.5)	12.144	0.902	1.791	1.864	1.942	0.002	3.285	4.192	4.983
(1,000,log n/n,3.5)	14.360	1.389	1.710	1.764	1.839	0.002	12.175	13.806	15.060
(1,000,0.1,3.5)	15.053	1.442	1.681	1.736	1.819	0.002	12.816	14.145	15.525

from the previous period. Specifically, given $\mu(t-1)$, we solve the following optimization problem at each time step t :

$$\begin{aligned} & \underset{\mu(t)}{\text{maximize}} \quad \sum_{i \in V} \frac{1}{\gamma} \left(v_i + \beta \sum_{j \in \mathcal{N}_i} \frac{\mu_j(t-1)}{d_i} + \ln \frac{1 - \mu_i(t)}{\mu_i(t)} \right) \mu_i(t) \\ & \text{subject to} \quad 0 \leq \mu_i(t) \leq 1, \forall i \in V. \end{aligned}$$

We still use the real-world network—*Amherst41* as the numerical illustration, similar to the experiment in Appendix E.5. We set the price sensitivity to $\gamma = 0.1$, the intrinsic value $v_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{U}(0, 4)$, and the network effect intensity $\beta = 3$ in order to satisfy the condition in Theorem 6. In Figures EC.14a and EC.14c, we present the averages of optimal dynamic prices and adoption probabilities across all agents, respectively. The results show that the optimal pricing strategy involves using price as a leverage tool to quickly drive agents' behavior toward the optimal static case within the first few time steps (approximately 3 steps), after which the price stabilizes at the optimal static level. Despite the negative impact that price should typically have on adoption, the existence of network effect causes both price and adoption probability to rise in the early stages under the optimal pricing strategy. Similarly, the myopic pricing scheme also sees an initial increase in both price and adoption, but its stabilized price is slightly higher than the optimal static price, resulting in suboptimal long-term behavior. In Figures EC.14b and EC.14d, we display the price and adoption probability trajectories for 10 randomly selected agents from a specific network instance. Under perfect price discrimination, different agents experience distinct prices, but their price trajectories follow a pattern similar to the average price. This dynamic pricing pattern suggests that, with network effects, the platform first uses a low price to attract more customers and then stabilizes at the optimal static price to maximize long-term profit.

Finally, in Figure EC.15, we demonstrate the performance of different pricing schemes. Both the myopic and dynamic schemes yield higher profits than the static scheme during the early time steps (i.e., the transient phase). However, as time progresses, these gains diminish. In the long run, the dynamic scheme converges to the same performance as the static scheme, while the myopic scheme remains slightly lower.

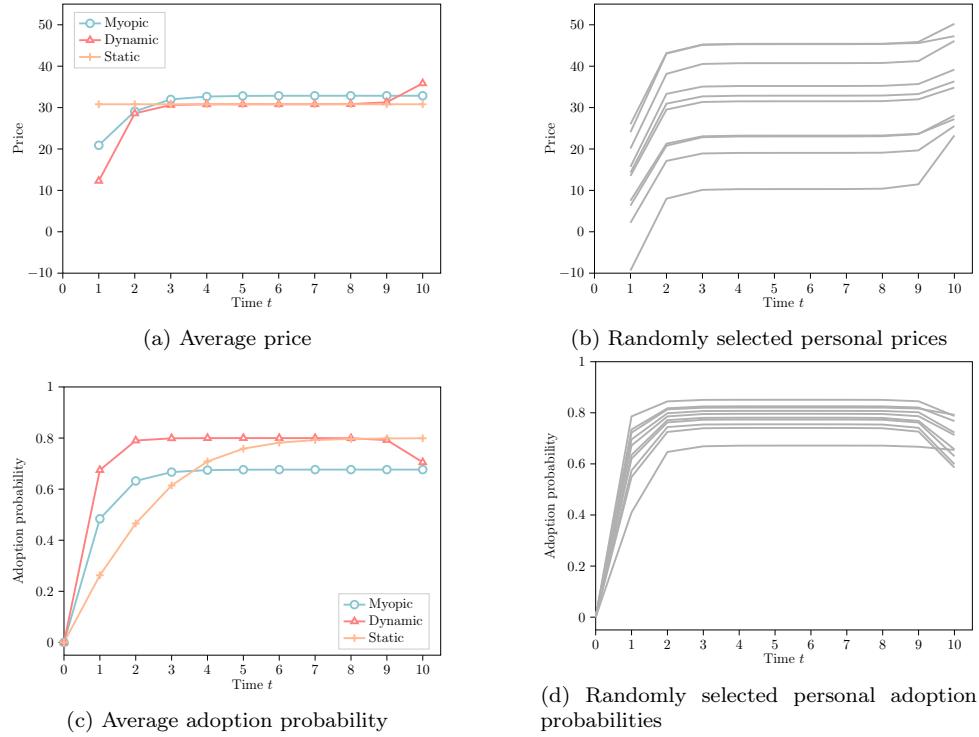


Figure EC.14 Different dynamic pricing schemes under the perfect price discrimination.

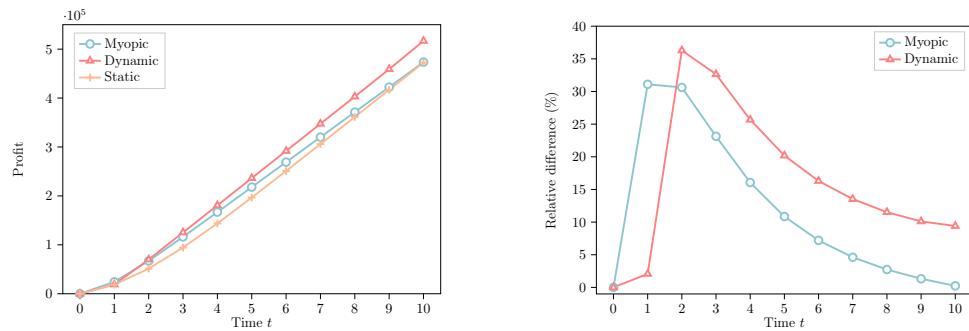


Figure EC.15 The performance of different dynamic pricing schemes under the perfect price discrimination. Left: Accumulative profit; Right: Relative difference in profit where Static is set as the benchmark.